PRESIDENCY UNIVERSITY
KOLKATA

Report on Departmental Pre-Submission Seminar for Ph.D. Programme

1	Name of the Candidate	Chayan Bhattacharjee
2	Department	Department of Life Sciences
3	Title of Thesis (As per Ph.D. Registration Certificate)	**Investigating the molecular details of Hepatitis C virus entry events**
4	Date of receipt of 5000-word summary by the Convener	28th November, 2020
5	Date of Seminar Presentation	23rd December, 2020

Signature of the Supervisor

Date: 23/12/2020

Signature of the Co-supervisor (if any)

Date:

This is to certify that (name of the candidate) presented his/her work in a Departmental Seminar in defense of his/her thesis in presence of Ph.D. Committee, Faculty members and Research/MA/MSc/BA/BSc students.

(Signature of the Convener & Members of the Departmental Ph.D. Committee)

1. Prof.Anita Mukherjee

2.Prof.Maitree Bhattacharyya

3. Prof. Prabir Kumar Mukhopadhyay

4. Prof. Devashish Sen

5. Dr. Malay Das

6. Dr. Aparna Mukhopadhyay

7. Dr. Sanghamitra Dey

8. Dr. Santanu Chakraborty

9. Dr. Puja Ray

Report on the performance of candidate: May be allowed to submit thesis as per University guidelines.

Signature of the Convener of the Departmental Ph.D. Committee

Title

Investigating the molecular details of Hepatitis C virus entry events

Name: Chayan Bhattacharjee

Registration number: R-14RS18110016

Name of supervisor: Dr. Aparna Mukhopadhyay

Name: Chayan Bhattacharjee

Supervisor: Dr. Aparna Mukhopadhyay

Title of Thesis: Investigating the molecular details of Hepatitis C virus entry events.

PhD Registration Number with date: R-14RS18110016

Enrolment Number with date: Sc/DOC/13-14/11

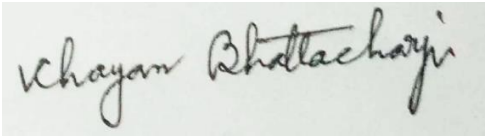
Place of work: Presidency University

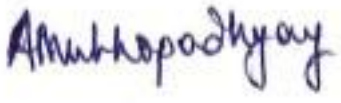
Residential Address: 124, KCC Sarani, Nazrul Park, Po. Bhadreswar, Dist. Hooghly, Pin. 712124.

Corresponding Address: 124, KCC Sarani, Nazrul Park, Po. Bhadreswar, Dist. Hooghly, Pin. 712124.

Email: bhattacharjee.biotech@gmail.com, chayan.rs@presiuniv.ac.in

Phone Number: 9007475328





Signature of Research Scholar

Dated: 12-12-2020

Signature of Supervisor

Dated: 12-12-2020

Contents

Sl No.	Subject	Page number
1.	Abstract	4-5
2.	Introduction	6-7
3.	Objectives (as per approved in 1000 word)	7
4.	Methodology	7 - 13
5.	Results (objective wise)	13- 21
6.	Discussion and Future Prospects	21-22
7.	References	23-25
8.	Tables and Figures	26-66
9.	Appendix	66-69
10.	Publication	70
11.	Certificates from Conferences Attended	71-73
12.	Plagiarism Certificate	74

1. Abstract

The positive stranded RNA genome containing, enveloped virus Hepatitis C (HCV) is the leading cause for chronic liver disease and hepatocellular carcinoma. To understand differences in HCV entry due to genotypic variation, 29 nucleotide sequences of the envelope glycoprotein E2 from two HCV genotypes and their subtypes (specifically 1a, 1b, 3a and 3b) were analysed. Nucleotide analysis revealed a preference for C- richness which was attributed to a bias for Cytosine rich codon in E2 of genotype 1, a similar but lower preference in genotype 3. Comparison of amino acids revealed a conserved N-terminal region where most of viral antigenic sites have been found, flanked by the hypervariable regions, with majority of the inter-genotypic changes concentrated at the C-terminal portion, a region mostly involved in host receptor attachment. Modelling followed by docking analysis between de novo models of E2 glycoproteins and PDB derived host receptor protein CD81 revealed unique interacting residues for each individual genotype in both HCV E2 and the CD81 with genotype 3a exhibiting the strongest interaction with CD81.

In our quest to understand the early events of HCV entry, fluorescent (GFP-tagged pseudoparticles were generated which were used for various trafficking assays via confocal imaging. We observed internalization and recycling of fluorescence within 15 minutes of internalisation by quantitation at various time points and co-localisation with plasma membrane marker FM 4-64. From our observations we predict that fusion of viral and cellular membranes occur within 15 minutes of internalisation. To understand the role of regulatory GTPases, particularly Rab1a, similar experiments were repeated in a Rab1a knockdown (Rab1aKD) cell line. In these cells we observed impaired trafficking of fluorescent HCVpp indicating an important role of Rab1a in HCV internalisation.

To identify entry inhibitors of HCVpp from naturally occurring medicinal plants, methanolic leaf extracts were prepared from Guava (*Psidium guajava*), Champa (*Magnolia champaca*), Black plum (*Syzygium cumini*) and Tamarind (*Tamarindus indica*) plants known to possess antiviral and hepatoprotective property. To explain the antiviral activity among these four leaf extracts, we observed only methanolic leaf extract of Tamarind to exhibit an inhibitory role ($p < 0.05$) by both qPCR and microscopic experiments.

In conclusion, we have established stark differences in structure and binding activities of the glycoprotein E2 across genotypes 1 and 3, information that needs to be kept in mind while prescribing therapeutics tested against genotype 1 strains. We have also established a

trafficking assay using fluorescent HCVpp by which we have been able to establish a probable time line of entry. We have also reported a role of Rab1a in HCV trafficking. This information will enable researchers to better understand the biology of the virus and design newer preventives that could be interfering with the entry process. To establish a potent natural HCV therapeutics, we found Tamarind leaf extract to act as a potent entry inhibitor for HCV. However further mechanistic details still need to be investigated.

2. Introduction

Hepatitis C virus (HCV) infection is considered to be a global health problem, with more than 70 million infected people worldwide[1, 2] and the disease has 1- 1.9% prevalence in India[3]. A primarily asymptomatic acute infection often becomes chronic, causing liver disease ranging from fibrosis, cirrhosis to hepatocellular carcinoma and ultimately death[4, 5]. Transfusion transmitted infection (TTI) remains a major concern especially for thalassemia patients in India[6]. Treatment is restricted to a few drugs that are expensive, narrow genotypic coverage and multiple side effects[7]. Hence development of better therapeutics and preventives require a perfect understanding of the HCV biology. Furthermore, the massive variability of this positive-stranded RNA virus makes it difficult to manufacture effective therapeutic vaccines to prevent the pathogen from spreading. Infection models have remained limited to chimpanzee[8], immunodeficient mice[9] or presently treeshrew[10]. Due to lack of proof reading activity of viral RNA dependent RNA polymerase (RdRp), a lot of variation exists within the genome which gives rise to eight HCV genotypes and millions of quasispecies. Among all genotypes, genotype 1 and 3 have been found to be predominant in the Indian subcontinent, while infection of genotype 1 is most prevalent worldwide[11, 12] .

The 10kb genome of the virus has a large open reading frame, encoding 11 structural and nonstructural proteins including, a core protein, two glycoproteins E1 and E2 and a number of non-structural proteins[13]. After entering the liver cells from sinusoidal blood, HCV interacts with several cellular factors and receptors (including the tetraspanin receptor protein CD81) that trigger virus uptake into the hepatocytes. Viral entry proceeds as attachment followed by clathrin mediated endocytosis and fusion between viral and cellular membranes to release the viral genome for establishment of infection. At the time of entry, several hepatocellular receptors like SRBI, LDL receptors, CLD1 and most importantly tetraspanin receptor CD81 bind with HCV glycoprotein E1E2, which are masked with different lipid molecules forming HCV lipo-viral-particle (HCV-LVP)[14]. E2 is a better characterized subunit than E1 and is believed to be subunit for HCV entry and endocytosis. HCV pseudoparticles (HCVpp, non-infectious lentiviral particles expressing HCV envelope proteins)[15, 16] and cell-culture derived HCV (HCVcc, a strain of HCV capable of being propagated in cell culture)[17,18], has allowed the study of entry in detail. During clathrin coated vesicle endocytosis, multiple fission-fusion and sorting events results in the generation of either recycling endosomes destined for the cell surface or late endosomes that ultimately fuse with lysosomes for degradation[19]. Regulation of such sorting is not completely

understood but is believed to be mediated by the small Rab (Ras-related in brain) GTPases, particularly that of Rab1a, Rab4, Rab5, Rab 7 and Rab11 along with an intact microtubular network[20,21,22]. In case of HCV endocytosis, a role of Rab5 and Rab13[23,24] has been reported. However, Rab1a is an established regulator for endocytic sorting for many ligands but has not yet been shown to play a role in HCV entry.

Medicinal plants and their natural compounds have been used as traditional medicine for liver related diseases and antiviral therapeutics all over the world[25,26,27]. However their mode of action, inhibitory property and efficacy are not known very well. Guava, Champa, Black plum and Tamarind are common plants in India, especially in Kolkata for their ethnomedicinal usage. Various parts of these plants, such as seeds, leaf, roots, bark have been extensively used for treating fever, constipation, indigestion analgesic substance in traditional India[28,29,30,31] .

3. Objectives

1. Compare variations of glycoprotein E2 across genotypes and determine differences in receptor interaction *in silico*.
2. Study cellular factors that regulate the HCV entry events, particularly the role of Rab GTPases.
3. Screen natural compounds predicted to improve the prognosis of HCV infections to study their properties to impair entry.

4. Methodology:

4.1: Selection of HCV sequences

E2 glycoprotein of HCV strains 1a, 1b, 3a and 3b were retrieved from online protein databases as mentioned in Table 1. A total of 29 sequences of HCV genotype 1 and 3 were selected for this study according to availability of whole genome sequences in the database, all listed in Table 2. To identify any intragenotypic differences of four nucleotides among 29 sequences, t test analysis was conducted. Similarly amino acid sequences of glycoprotein

HCV E1, RNA dependent RNA polymerase (RdRp) and HCV whole genome were retrieved from NCBI database.

4.2: Nucleotide compositional, frequency analysis, codon usage and cluster codon analysis of HCV E2

Nucleotide composition of E2, whole genome, E1 and HCV RdRp and usage of each codon and their frequency were analysed by online tools (Table 1). Differentiation of amino acids character such as codon usage, differences in intragenic and intergenetic variation was analysed manually (Table 1).

4.3: Position wise chemical property

Position wise characteristic differences were determined for all the amino acids present in E2 paired between strains 1a, 1b, 3a and 3b. For this all amino acids and their respective codons were paired and any change in amino acid between strains and genotypes, were colour coded and categorized.

4.3: *In silico* Protein Modelling of E2 protein and receptor protein CD81

Tools used to characterise E2 and CD81 are listed in Table 1. Due to unavailability of template structures in Protein Data Bank (PDB), the three-dimensional structure of E2 was generated by template-based modelling through several online modelling servers (Table 1). The best models of each server were protonated at ionic strength, physiological pH and subjected to loop refinement. Secondary structure assignments of best models of each modelling server were compared with truncated models existing in PDB. Energy minimisation of each model were minimised by Smart minimizer of Discovery studio (DS 2.5) to satisfy a RMS (Root Mean Square) gradient of 0.1kcal/ molA^{o}. A theoretical structure of CD81 available in PDB was also considered in our study. The quality of energy minimised structures were further validated through different online protein structure validation tools in SAVES V5.0 server (http://services.mbi.ucla.edu/SAVES/). To establish the authenticity of our work, the models were compared with published literature, relative solvent accessibility

(RSA) an indicator of percentage of residue accessible to solvent was determined. Residues with RSA more than 16% was considered to be exposed residues[32].

4.4: Differential *in silico* interactions of cellular protein CD81 and glycoprotein E2 of genotype 1 and 3

To determine cellular binding site of E2, structure and sequence based prediction servers were utilised (Table 1). Considering the variance in CD81 binding site of E2, ZDOCK rigid body in Discovery Studio 2.5 was employed to identify the exact binding site residues of HCV genotype 1a, 1b, 3a and 3b against the known binding residues (Leu162, Thr163, Ile181, Ile182, Ser183, Asn184, Leu185, Phe186, Lys187) of CD81[33, 34], without defining the CD81 binding site of E2 glycoproteins. The possible C terminal transmembrane portion present in E2 was excluded. The CD81-E2 binding interface residues of 100 RDOC poses was considered to be the CD81 binding site of E2 according to their appearance and frequency in receptor-E2 binding interface. To increase specificity, we conducted another E2-CD81 docking through ZDOCK with defined binding site residues collected from previous docking process with four genotypes. The best poses of each complex from all four genotype were chosen by RDOCK score and analysed by different biophysical properties like interaction energy (IE), solvent accessible surface area (SASA), hydrogen bond (HB) and binding free energy in DS 2.5.

4.5: Chemicals and reagents

Major chemicals, reagents, cell lines, plasmids and primers used in our study are mentioned in Appendix 1.

4.6: E1E2 -GFP fusion construct preparation

E1E2 sequence was amplified from pcDNA E1E2 vector using two primers (Table 25), After amplification and subsequent digestion with *Hind*III and *Eco*RI (NEB, USA), E1E2 region was ligated into pEGFP-N1 to create a E1E1-EGFP fusion construct with GFP at the C-terminal end. Colonies were picked and screened.

4.7: Cell surface expression of EGFP

HEK 293T cells were seeded on sterile coverslips to achieve 70-80% confluency. pEGFP-E1E2 vector was transfected using TurboFect (Thermo Fisher Scientific) reagent following manufacturer's instructions. 36h post transfection cells were stained with 5µg/ml of plasma membrane marker FM 4-64 for 15min on ice followed by fixation with 4% formaldehyde solution for 10min on ice and 10min at room temperature. Overexpression of E1E2-EGFP was checked by confocal microscopy (Leica DMi8) using 63X oil immersion objective.

4.8: Western blot analysis

To confirm protein expression of pE1E2-EGFP in HEK293T cells, western blot analysis was performed. 48h post transfection, HEK283T cells were scraped and lysed in 100µl lysis buffer [50mM Tris, 150mM NaCl, and 0.1% NP40 supplemented with 100mM PMSF (Themo Fisher Scientific) and protease inhibitor (Sigma)]. Lysed cells were clarified at 16,000g for 20min. 30µg protein was separated in 12% SDS-PAGE, followed by immunoblotting with anti GFP antibody.

4.9: Production of fluorescent HCV pseudoparticles (HCVpp)

HCV pseudoparticles were produced by a modification of published procedure[35]. Cells were seeded on 6cm dishes a day before transfection to achieve 50 – 80% confluency. Constructs, HCV E1E2-EGFP (HCV glycoprotein E1E2 fused to GFP), retroviral packaging construct pTG5349 and a reporter pLenti-PGK-GFP were transfected into HEK 293T cells by calcium-phosphate method (Calphos Mammalian Transfection kit, Clontech) as per manufacturer's instructions. pTG5349/gag-pol (4.2µg), pLenti-PKG-GFP (4.2µg) and pE1E2-EGFP (1.4µg) plasmid DNA were used. The transfection mixture was added to HEK 239T cells and kept at 37ºC in 5% CO_2. 40h post transfection, HCVpp containing media (approximately 42 ml) was collected, purified via passage through 0.45µm filter (Millipore, MILLEX-GV), concentrated by Amicon ultra filter (Amicon Ultra Centrifugal Filters, Merck) and finally stored at -80ºC. Approximately 900µl HCVpp was obtained from 42ml of media.

4.10: Validation of HCVpp with qRt-PCR

Confluent Huh7 cells monolayers grown in 10cm plate were transduced with 50µl HCVpp supplemented with DMEM devoid of FBS and antibiotics on ice for 1h. This plate was

incubated for 20 min in followed by total RNA extraction using Trizol reagent (Sisco research laboratories, India). cDNA was synthesized with 1µg of RNA, using iScript^{TM} Reverse Transcription supermix for RT-qPCR kit (BIO-RAD Laboratories, USA) following manufacturer's instructions at PCR condition (25ºC for 5min, 46ºC x 20min , 95ºC x 1min at RT (25°C)). 60ng of cDNA was amplified by PCR using 20pM/ µl of primers for PKG-GFP. GAPDH was used as a housekeeping gene for every PCR reaction. All the primer sequences are presented in (Appendix 1). The PCR amplification was performed as per the following cycle: 3min x 95ºC, 34 cycles of 95ºCx30sec, 47ºCx30sec and 72ºCx1min at. The PCR amplified product was checked by 1.5% agarose gel electrophoresis.

4.11: Study of entry kinetics and cellular trafficking of HCVpp in Huh7 and Rab1a KD cell line

Approximately 1.04×10^6 cells (Huh7, Rab1aKD) were seeded on sterile coverslips (designated as 0, 5, 10, 15 and 30 min time points). 50µl HCVpp diluted in 450µl of DMEM (without FBS, antibiotics) were added dropwise on the on coverslips (100µl mixture each) on ice. After washes, HCVpp was allowed to internalise by shifting cells to 37ºC time periods as indicated. Internalisation was stopped by shifting the cells to ice, followed by fixation in 4% formaldehyde for 10min on ice and 10min at room temperature (RT). For staining the plasma membrane 5µg/ml FM 4-64 was incubated for 15min on ice prior to fixation. Similarly, to study lysosomal colocalisation, lysotracker was added at a concentration of 0.14µM (diluted in PBS) to both of the cell lines for 30min at RT. Cells were examined via confocal microscope (Leica DMi8) in 63X oil immersion and images quantified. Microscopic images were quantified by ImageJ-Fiji software (ImageJ-win-64). Images were pseudo-coloured and co-localisation measured using Image J.

4.12: Preparation of plant Extracts

Guava, Champa, Black Plum and Tamarind methanolic leaf extracts were prepared by cold extraction method, where the followings solvents (n-hexane, chloroform and methanol) were used according to their increasing polarity. The plant leaves were collected from various parts of Kolkata between the months of March to April. Healthy leaves were selected and air-dried for at least seven days at room temperature. The dried leaves were crushed and powdered. About 50g of powder from each plant was subjected to solvent gradient extraction process.

N-hexane was added to individual leaf powders and incubated in room temperature (24ºC) at 120rpm for 24h. The solvent with dissolved extracted fraction was decanted out and collected and chloroform added to the residue and process repeated. After 24h of incubation at 120rpm, extracted solutes were decanted followed by methanol to obtain the methanolic solvent fraction. Each of the collected fractions were further concentrated in a rotary evaporator to obtain a gelatinous substance. All the extracts were dissolved in dimethyl sulfoxide (DMSO) (Sisco Research Laboratories Pvt. Ltd., India) at 100mg/ml concentration. DMSO dissolved extracts were passed through 0.22µm filter and stored at -20ºC.

4.13: Cell viability assay

To investigate cell viability MTT assay was done following the protocol of Pandey et al. with some modification[36]. In brief, 3×10^4 cells/well was plated into 96-well plates (Thermo fisher scientific). Various concentrations (50 to 1000µg/ml) of methanolic leaf extracts was added to the cells incubated at 37°C for 48h. DMSO was used as control. Post-treatment, the media was removed from each well 5 mg/ml 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide (MTT) solution was added to each well. 3 hours later media was carefully removed and 100µl of DMSO was added to dissolve the formazan crystals and luminescence measured with a microplate reader (Synergy/ H1, Biotek, USA) at 540 nm and a reference wavelength of 620nm.

4.14: Effects of methanolic leaf extracts on HCV entry

Huh7 cells were seeded, transduced with HCVpp followed by exposure to the different plant extracts same as previously mentioned for MTT assay. After 1h, cells were washed and incubated at 37°C for 20min. For HCVpp entry and endocytosis assay, the investigation was divided into two categories, (A) determination of KGFP expression by real-time quantitative PCR (qPCR) and (B) estimation of EGFP (E1E2-EGFP) expression and plasma membrane co-localisation with FM 4-64 was examined via confocal microscope and quantified by ImageJ-Fiji software (same as previously mentioned).

4.15: Statistical analysis

All statistical analyses were performed using SPSS 20.0.and Microsoft Excel (Microsoft Excel 2010). The type of analysis performed for each experiment is mentioned in the table footnotes/ figure legends.

5. Results:

Objective I: Compare variations of glycoprotein E2 across genotypes and determine differences in receptor interaction *in silico.*

5.1.1: Nucleotide composition analysis shows a Cytosine bias across all the genotypes

To analyse differences, 29 sequences of HCV of genotypes 1a, 1b, 3a and 3b were taken (Table 2). Nucleotide analysis of 29 sequences of E2 clearly indicate that among the four genotypes, the most frequent ($p< 0.001$) nucleotide was cytosine (C), followed by guanine (G), thymine (T) and adenine (A). Genotype 1a exhibited the highest average base percentage of Cytosine (C) followed by the other genotypes (Table 3). However, C content did not vary significantly within a particular genotype but varied significantly ($p< 0.001$) between the two genotypes (Table 3).

On analysing the preference of C exhibited by any other HCV protein, glycoprotein E1 and the more conserved HCV protein RNA dependent RNA polymerase (RdRp) and entire genome (Table 2), almost similar results were obtained. We observed that the entire HCV genome (Table 4) and RdRp (Table 6) from all strains revealed a distinct preference for C (a significant difference of C with other nucleotide present in an individual genome, $p<0$ 0.05), similar to that seen in HCV E2 (Table 3). However, Glycoprotein E1 differed slightly and exhibited almost similar preference for both Cytosine and Guanosine in all strains (Table 5). Avoidance of Adenine in E1, RdRP and whole genome was common for all strains ($p <$ 0.05) which support the previous hypothesis that RNA viruses being GC rich.

5.1.2: Nucleotide position analysis reveals a preference of C-rich terminal codon

Position wise nucleotide analysis of HCV glycoprotein E2 revealed a preference for C at 3rd position of any codon compared to other nucleotides for both the genotypes 1 and 3 (Table 7). Genotype 1 (average percentage of genotype 1a and 1b) in general had greater preference for C in the 3rd nucleotide position than genotype 3, exhibiting a 42% occurrence of cytosine

for genotype 1 and 33% occurrence for genotype. Moreover, one-way ANOVA followed by Bonferonni post hoc analysis showed that except 1b the C content both in the 1st and 2nd position of 1a, 3a and 3b didn't vary significantly ($p > 0.001$, SD=1.65 to 2.88) from the other nucleotides. C content for all genotypes varied significantly ($p < 0.001$) at the 3rd position. However, the least preferred nucleotide for 3rd position of any codon for any amino acids were either A or T.

A similar trend was observed with HCV glycoprotein E1 (Table 9) and polymerase RdRp (Table 10). However, the entire genome revealed something different (Table 8), G richness in the 3rd position instead of C of genotype 1a. The other genotypes (1b, 3a and 3b) a significantly ($p<0.01$) greater preference for C to other nucleotides within its genome, similar to our previous results of E2 was observed.

5.1.3: Cluster codon analyses reveal preference of C rich codons across most of the E2 sequences

Cluster codon analysis, results clearly indicate a preference for those codons that had a terminal nucleotide C. For example, out of two codons of aspartate, GAT and GAC, all genotypes had a preference for GAC (63.4%). However some exceptions were also observed in Glutamine, Serine and Aspartic acid, genotype 1 preferred a C-rich codon, but genotype 3 didn't (Table 11). Those amino acids with four or more codons, preferred codons with either a G or a C in the terminal position, with avoidance for those with an A or a T. This pattern was observed in the codon usage of Gly and Leu. Arg and Val codons ending with G (26% and 50% respectively) were preferred over those ending with C (25% and 29% respectively) in all HCV-E2. For few amino acids like Ala, Asp, Glu, His, Ile, Lys, Gln and Ser, no clear preference was observed. However, codon like Met (ATG) and Trp (TGG) also preferred G rich codon where they had only one codon. Moreover, most GC rich codon among all amino acids for E2 was Gly (GGC), Pro (CCC), Arg (CGG) and Val (GTG).

5.1.4: Amino acids present in genotype 1 is more conserved compared to genotype 3 in-E2

As our previous results indicate a genotype specific difference between 1 and 3 with codon preferences, we decided to extend our analysis to determine functional differences. NCBI-BLAST analysis and multiple sequence alignment or CLUSTALW analysis (EMBL-EBI in

Hinxton, Cambridge) of E2 among 29 amino acids sequences indicate some position wise differences between 1a and 1b (19% difference); 3a and 3b; (20% difference) and genotypes 1 and 3 (70% difference between 3b and 1b). These results suggested examining the nature of these differences in more detail. In these analysis amino acids sequences of 1a, 1b, 3a and 3b were paired and characterised in every amino acid position. As this work was tedious to compare all 29 sequences, we decided to take a representative strain randomly from each genotype (Table 12). t-test analysis showed there was no significant difference ($p > 0.05$) between our chosen strain and other strains of that genotype except for T and G content in a strain of 1b (Accession No. EU155366.2). Simple BLAST analysis revealed that all 1b strain shared 93% identity with EU155366.2. So, we decide to select EU155366.2 as a representative strain for 1b as earlier reports suggested despite sequence dissimilarity by about 30%, structural conservation is still maintained[37].

Alignment and scoring amino acid changes with colour codes are shown in Fig 1. It was observed that C terminal of HCV E2 protein showed maximum variation between genotypes 1 and 3 but this region was more conserved in genotype 1 compared to genotype 3. For genotype 1, both the subtypes (1a, 1b) and 3(3a, 3b) were conserved in their particular genotype but genotype 1 differed 47% with genotype 3. The N terminal region of E2 exhibited only 13% unchanged amino acids across all genotypes, (Fig 1B). Between N and C terminal, a conserved region was observed in the central portion of the protein, which meant conserved amino acids within genotype 1 and 3 but changes when 1 is compared to 3. Scoring changes, most of the amino acids changed from polar to non-polar. This type of change was seen largely between the genotypes 1 and 3. However, within the genotype the changes were less profound. Minimal changes in the amino acids of E2 between genotype 1 and 3 were changes from basic amino acids to acidic and vice versa. Changes spanned almost the entire length of the E2 protein but amino acids from 406 to 411 regions remained conserved among the two genotypes. While the biological significance of this region is still not clear, previous literature state that residues 412-423 represent the most conserved antigenic site of E2. These findings called for more comprehensive studies to understand how this variations effects protein biology between genotype 1 and 3.

In a similar experiment we calculated position of nucleotide alterations that differed between strains and genotypes for each codon (Table. 13). The analysis showed maximum changes were tolerated in terminal nucleotide of any codon (12% in genotype 1, 7% in genotype 3). However, genotype 1 tolerated fewer changes in other position (1^{st} or 2^{nd}

nucleotide) of any codon compare to genotype 3. This finding again proved our previous analysis that genotype 1 contained more conserved amino acids compared to genotype 3.

5.1.5: De novo modelling of HCV glycoprotein E2 and cellular receptor CD 81

According to our analysis, it was evident that between HCV genotypes and their subtypes, nucleotide to amino acids variations were present which has the possibility to induce a biological effect by causing alterations in the protein structure and thereby protein function. In such a case, due to unavailability of any complete crystal structure of E2 in protein data bank (PDB) or any other published literature, de novo modelling was done by different online *in silico* protein modelling tools (Table 1) including MODELLER software[38]. This work was done in collaboration with Dr. Suman Nandy, Tata medical center, Kolkata. The best models for each of the representative strain of 1a, 1b, 3a and 3b were chosen post energy minimisation (Table 14, Figure 2). Comparing our selected de novo models with several published incomplete models and structural information of E2, we found a strong similarity of our models with others [39] [40]. As an additional validation, those residues known to be exposed and antigenic determinants from published literature were compared to solvent accessible residues of our generated models (Table 15). For receptor protein CD81, the complete structure was available in PDB (PDB ID: 2AVZ), which was also predicted similarly as HCV E2 and validated. Post validation we used energy minimised 2AVZ structure of CD81 for our molecular docking experiments.

5.1.6: Molecular interaction of CD81 and HCV E2

Depending on molecular docking results, it was observed that the CD81 binding region of E2 for the both genotypes and their subtypes was unique and it spanned from 470 to 700 amino acids position (Fig 3) with the specific residues listed in Table 16. Between E2 of genotypes 1 and 3, only a few residues were common. Glycoprotein E2 of genotype 3 (3a and 3b) did not have any common binding residues. Within CD81, only two residues were detected to be commonly involved in interaction with E2 from all genotypes (Table 16). The result clearly showed that even the residues of the host cellular receptor protein CD81 did not remain conserved while interacting with E2 from different genotypes or strains. E2 and receptor protein CD81 binding affinity was mostly dominated by hydrophobic interactions, electrostatic interactions, and numerous hydrogen bonds. Solvent accessible surface area,

SASA>2000 Å2 was determined for each complex. The results indicated that, genotype 3 has higher affinity with receptor protein CD81 with greater binding energy (Table 17). However, genotype 3a had highest binding affinity with CD81 among the four HCV strains. In this interaction, CD81 depicted four interaction patches that contributed an electrostatic interaction, involving interacting residues like Ser160, Thr163, Asp117, Thr167, Ala164, Glu188, Asp189, Asp196, Luc 165 and Asp195. However, Van der Waals (VDW) contacts contributed most in E2-CD81 complex for HCV genotype 3a. Taken together genotype 3a showed strongest interaction with CD81 with its independent binding residues.

Objective II: Study cellular factors that regulate the HCV entry events, particularly the role of Rab GTPases.

5.2.1 pE1E2-EGFP construct successfully expressed EGFP protein in 293T cells

E1E2-EGFP fusion protein was conceptualized in order to generate fluorescent pseudoparticles suitable for microscopy studies. A C-terminal GFP tagged fusion of E1E2 was created such that ER-signal sequence of the protein was undisturbed. Furthermore as the C-terminal end of E2 is on the cytoplasmic side, a GFP tag is expected to not significantly affect its ability to interact with the receptor protein CD81.

As detailed in material and methods and depicted schematically in Fig 4A, E1E2 was amplified by PCR and inserted into pEGFPN1 to create the E1E2-EGFP fusion construct. Insertion of E1E2 was checked before proceeding further (Fig 4B). To check whether the construct was successfully prepared, protein expression of EGFP was analyzed. 48h of post transfection EGFP expression was analyzed by immunofluorescence analysis (Fig. 5 A, B) as well as by immunoblotting (Fig 5C). GFP expression was observed to be in vesicular structures throughout the cytoplasm. Quantitation of the pixel intensity revealed significantly higher fluorescence intensity signal compared to mock transfected cells (Fig 5B).

5.2.2: Preparation and validation of fluorescent pseudoparticles

Pseudoparticles (HCVpp) were prepared as depicted schematically in Fig 6 and detailed in the methods section. Briefly, transfection of the E1E1-EGFP, a Gag/Pol packaging construct and a GFP or Luciferase containing reporter construct into 293T cells yielded fluorescent pseudoparticles that could be purified from the media. To check for successful production of

HCVPpp in HEK 293T cells, harvested HCVpps were allowed to internalize for 20min in Huh7 cell line. 20min post internalization, PKG-GFP expression was detected by qRT-PCR (Fig 7B) from isolated total RNA of the transduced cells which indicated successful production of pseudoparticles. To determine EGFP protein expression in E1E2 glycoprotein of HCVpp, we transduced Huh7 cells grown on coverslips with HCVpp and fixed them after 20 min. EGFP expression was observed under confocal microscope (Fig 7A). The transduction efficiency was approximately 70% (Table 18).

5.2.3: HCVpp Trafficking in Huh7

To study HCVpp trafficking, we conducted a time lapse experiment of HCVpp in Huh7 cells. Fig 8 shows representative images of Huh7 cells that were fixed 0, 5, 10, 15 and 30min after HCVpp internalisation. We observed an almost steady expression of EGFP at earlier times (up to 15min) and then a gradual decline by 30min when compared to early time points.

Measurement of pixel intensities of EGFP tagged HCVpp in Huh7 is depicted in Fig 9B (blue bars). It was observed that while pixel intensity remained constant between 0- 5 minutes, pixel intensity rose to a peak level by about 15 minutes. This is possibly due to coalescing of vesicles with endocytic processing. Thereafter the pixel intensity dropped by about 30 minutes of internalisation possibly due to degradation of E1E2-EGFP. From these results, we hypothesized that fusion between membrane and viral membranes occur between 5- 30 minutes.

To further study the trafficking details, the cells were stained with a plasma membrane dye FM4-64. Fig 9A shows representative individual and merged images and Fig 9C shows colocalisation between HCVpp and FM4-64 at various time points. As seen in the images and quantitative data, colocalisation of HCVpp and FM4-64 rose sharply to a peak value at 15 minutes. 15 minutes was also the time when maximum pixel intensity of HCVpp was being observed (Fig 9B). As FM4-64 is a plasma membrane dye that is not internalised, a high colocalisation at 15 min (Fig 9C) is indicative of trafficking of E1E2-EGFP to the plasma membrane. This would indicate that viral and host membrane fusion has been completed by 15min of internalisation and hence the E1E1-EGFP has trafficking back to the cell surface via some recycling endosome. However by 30 minutes a drop in co localisation as well as E1E2-EGFP pixel intensity is perhaps indicative of either dissipation of E1E2-EGFP or its degradation. To cross check degradation in lysosomes, we used a lysosome marker

Lysotracker and conducted similar type of time lapse experiment. Interestingly we did not find any colocalisation even at 30min of internalisation in Huh7 (Fig. 10). This finding indicates that E1E2 is not degraded in lysosome after 30 min of internalisation.

5.2.4: Rab1a is important for HCVpp trafficking

In order to further understand the role of endocytosis, the effect of the RabGTPase, Rab1a in HCV trafficking was studied. Rab1a, apart from its role in ER-Golgi trafficking has also been previously characterised to be important in early endocytic sorting events as well as trafficking of early endosomes and their maturation into late endosomes by recruiting kinesin motor KifC2 thereby enabling microtubule based movement[22]. We decided to investigate the role of Rab1a in HCV trafficking by using a Rab1a KD cell line. Experiments previously described for Huh7 cells were performed in this cell line and compared to the events occurring in Huh7. Observation of trafficking and measurement of pixel intensity, (Fig 9B and C) revealed increasing pixel intensity over time. This was consistent with trafficking of other ligands in Rab1a KD cells lines, where accumulation of ligand had been observed. At all-time points, the pixel intensity of HCVpp in Rab1a KD cells was significantly higher than that of Huh7 cells (Fig 9B). This indicated a possible slower trafficking of HCVpp in Rab1a KD cell line, with higher pixel intensity indicative of coalescing vesicles. Colocalization with FM4-64 (Fig 9C) revealed that while Huh7 localisation dropped between 15- 30 minutes, there was no such drop observed in Rab1aKD cells. As Rab1a is known to slow down endocytic traffic at early time points, between 5- 15 minutes, we hypothesize that Rab1a has an important role to play in the endocytic trafficking of HCV and possible could also hamper fusion which we expect to be happening within 15 minutes of internalisation. To cross check for an association of E1E2-EGFP with lysosomes in the Rab1a KD cells, we co-stained with lysotracker and did not observe any co-localisation indicating a delay in trafficking at earlier time points (Fig 10). Even after 30min in Rab1a KD cells, E1E2-EGFP was possibly stuck in an earlier endocytic compartment and hence reaching lysosomes does not arise.

Objective III: Screen natural compounds predicted to improve the prognosis of HCV infections to study their properties to impair entry.

Four traditional medicinal plants (Guava, Champa, Black plum and Tamarind) were chosen to investigate their anti-viral properties against HCV entry (Table 19). These plants were chosen for their ethnomedicinal usage apart from being commonly available in Kolkata. Various parts of these plants, such as seeds, leaf, roots, bark have been extensively used for treating fever, constipation, indigestion and analgesic substance in traditional India. Several studies indicate the presence of different bioactive compounds that are responsible for many medicinal properties as listed in Table 20. Moreover leaf extracts of these plants have proven antiviral (anti Influenza virus), hepatoprotective and hypolipidemic activity becoming attractive agents to be tested against HCV.

5.3.1: Toxicity analysis of Methanolic leaf extract

Toxicity of the methanolic leaf extracts were tested in the given ranges: 200-700µg/ml for guava, 50-300µg/ml for champa, 50- 300µg/ml for black plum and 25-600µg/ml for tamarind. In case of guava extract, we observed a significant reduction ($p < 0.001$) of percent cell viability at a concentration of 400µg/ml. Similarly significant reduction ($p < 0.05$) in cell viability was observed at 200µg/ ml for champa extract, 75µg / ml for black plum ($p < 0.001$) and 250µg/ ml for tamarind ($p < 0.001$) (Fig 11). Optimised concentration for all the methanolic leaf extracts was selected where approximately 80% cell viability was observed and percent cell viability was not significantly reduced with their previous concentrations (Table 21). These optimised concentrations for each of the plant extract in respect to percent cell viability was selected and used in our further analysis.

5.3.2: Effect of methanolic leaf extracts on HCV entry

To determine the antiviral effect against HCV entry, Huh7 cells were infected with HCVpp in the presence of methanolic leaf extracts in their selected concentration. After 20 min of internalisation, GFP was estimated from RNA by qRT-PCR. We observed that only the methanolic leaf extract of tamarind exhibited a significant ($p< 0.05$) inhibitory effect on HCVpp attachment or entry to Huh7 cells at 100µg/ml concentration (Figure 12) in contrast to untreated control. Methanolic leaf extract of champa and black plum did not show any kind of inhibitory role in cellular attachment of HCV at their given concentration and conditions of the experiment.

Quantification of fluorescence intensity of EGFP after 20min of localisation we observed that consistent with earlier results, tamarind leaf extract treated cells (n= 99) exhibited a significant difference ($p< 0.001$) in EGFP fluorescence intensity (Figure 13 A, B). However, in this experiment, we also found a significant difference ($p< 0.05$) of EGFP intensity in cells treated with champa (n= 342) and guava (n= 519) leaf extracts, indicating reduced amount of E1E2-EGFP internalisation. Black plum leaf extract (n= 121) did not show any significant difference EGFP intensity consistent with our qPCR results. For further investigation, we conducted a colocalisation assay of a plasma membrane marker FM 4-64 with EGFP (Figure 14 A, B). In this case, we found no colocalisation of EGFP and FM 4-64 in cells treated with tamarind leaf extract (n= 17). However, champa and black plum extract treated cells showed higher percentage of HCVpp- EGFP co-localisation (55%, n= 20 and 53%, n= 17) than untreated cells (n= 20; 35%).

6. Discussion and Future Prospect:

In this study the three objectives have addressed distinct questions. In the first objective comparison of 29 HCV E2 sequences from genotypes 1a, 1b, 3a and 3b revealed previously unknown differences in nucleotide preferences, structure as well as interaction with the receptor protein CD81. Most interestingly these molecular variations were specific to individual strains/ genotypes, a fact that has been largely overlooked and could be an important determinant in the variable progression of disease. It is interesting to note that we observed genotype 3a to exhibit the strongest binding to CD81, a strain that is more virulent and prevalent in India. This also has immense implications in the quest to develop new HCV therapeutics or vaccines, especially for preventives as now more specific or pangenotypic tests need to be performed. Furthermore our understanding of viral biology is dependent on reagents prepared from primarily genotype 1a and 2a. Our results on receptor interaction calls for more in depth study in other HCV strains as well.

In our second objective we have described the preparation of fluorescent pseudoparticles. To the best of our knowledge this is the first description of fluorescent pseudoparticles which have been used in trafficking assays to determine the very early events of HCV entry. There are no studies looking at these very early time points. In our studies we have described the trafficking of HCVpp in Huh7 cells and predict approximate fusion timing. This information helps to understand the timing of viral entry and thereby design

methods of intervention. We also describe for the first time a role of Rab1a. These findings have potential to be tapped as a host targeting agent (HTA) to slow down HCV transmission and spread in infected individuals. Better understanding of the entry process helps in further research into preventives, an extremely important area of HCV research as many infections are spread via blood transfusions.

In our final objective we present data to show an inhibitory effect of tamarind leaf extract in HCV attachment to Huh7 cells. This could be via inhibition of HCVpp interaction with cellular receptors. Guava leaf extracts also exhibited some inhibitory effect. However, at higher concentration it could have better inhibitory effect like tamarind and this remains to be tested. There is a possibility that guava and champa extracts perhaps don't affect attachment (as evident from qRt-PCR data that includes PKG-GFP signal from those that were attached as well as those internalised) but may have an effect in HCVpp trafficking. This could explain the lower fluorescent intensity of E1E2-EGFP that has been observed. Overall, this preliminary data have a clear indication for future investigation to understand more about the mechanistic detail of the inhibitory effects of these plant extracts. Further studies would also involve testing with the active compounds present in these plant extracts.

Fig 15 depicts how our understanding of viral biology is enhanced by this present work. Our data reveal that viral fusion occurs between 5-15 minutes of internalisation and the envelope protein recycles back to the plasma membrane by 15 minutes. Thereafter it is degraded. The point of intervention by Tamarind extract has also been depicted. The role of Rab1a in early endocytic traffic of HCV is also shown.

Future work would involve laboratory validation of the *in silico* predictions to completely prove the genotypic differences in receptor interaction. Further in depth study of the mechanistic details of HCV entry, such as the role of other Rab proteins, the exact role of Rab1a in HCV entry, further time lapse studies to pinpoint the exact fusion time need to be done to clearly understand the events in HCV entry. The mechanism of inhibition of the plant extracts that show promise also need to be dissected in future studies.

References

1. Lavanchy D. The global burden of hepatitis C. In: *Liver International*. Vol 29. ; 2009. doi:10.1111/j.1478-3231.2008.01934.x

2. Alter MJ. Epidemiology of hepatitis C virus infection. *World J Gastroenterol*. 2007;13(17). doi:10.3748/wjg.v13.i17.2436

3. Saraswat V, Norris S, De Knegt RJ, et al. Historical epidemiology of hepatitis C virus (HCV) in select countries - Volume 2. *J Viral Hepat*. 2015;22(s1). doi:10.1111/jvh.12350

4. Koike K, Tsutsumi T, Miyoshi H, et al. Molecular basis for the synergy between alcohol and hepatitis C virus in hepatocarcinogenesis. In: *Journal of Gastroenterology and Hepatology (Australia)*. ; 2008. doi:10.1111/j.1440-1746.2007.05292.x

5. Willatt JM, Hussain HK, Adusumilli S, Marrero JA. MR Imaging of Hepatocellular Carcinoma in the Cirrhotic Liver: Challenges and Controversies. *Radiology*. 2008. doi:10.1148/radiol.2472061331

6. Goodnough LT. Risks of blood transfusion. *Anesthesiol Clin North America*. 2005. doi:10.1016/j.atc.2004.07.004

7. Pawlotsky JM. Treatment failure and resistance with direct-acting antiviral drugs against hepatitis C virus. *Hepatology*. 2011. doi:10.1002/hep.24262

8. Lanford RE, Bigger C. Advances in model systems for hepatitis C virus research. *Virology*. 2002. doi:10.1006/viro.2001.1316

9. Mercer DF, Schiller DE, Elliott JF, et al. Hepatitis C virus replication in mice with chimeric human livers. *Nat Med*. 2001. doi:10.1038/90968

10. Feng Y, Feng YM, Lu C, et al. Tree shrew, a potential animal model for hepatitis C, supports the infection and replication of HCV in vitro and in vivo. *J Gen Virol*. 2017. doi:10.1099/jgv.0.000869

11. Messina JP, Humphreys I, Flaxman A, et al. Global distribution and prevalence of hepatitis C virus genotypes. *Hepatology*. 2015. doi:10.1002/hep.27259

12. Uddin G, Shoeb D, Solaiman S, et al. Prevalence of chronic viral hepatitis in people of south Asian ethnicity living in England: The prevalence cannot necessarily be predicted from the prevalence in the country of origin. *J Viral Hepat*. 2010. doi:10.1111/j.1365-2893.2009.01240.x

13. Penin F, Dubuisson J, Rey FA, Moradpour D, Pawlotsky JM. Structural Biology of Hepatitis C Virus. *Hepatology*. 2004. doi:10.1002/hep.20032

14. Catanese MT, Uryu K, Kopp M, et al. Ultrastructural analysis of hepatitis C virus particles. *Proc Natl Acad Sci U S A*. 2013. doi:10.1073/pnas.1307527110

15. Bartosch B, Dubuisson J, Cosset F-L. Infectious Hepatitis C Virus Pseudo-particles Containing Functional E1–E2 Envelope Protein Complexes. *J Exp Med*. 2003. doi:10.1084/jem.20021756

16. Hsu M, Zhang J, Flint M, et al. Hepatitis C virus glycoproteins mediate pH-dependent cell entry of pseudotyped retroviral particles. *Proc Natl Acad Sci U S A*. 2003. doi:10.1073/pnas.0832180100

17. Wakita T, Pietschmann T, Kato T, et al. Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. *Nat Med*. 2005;11(7). doi:10.1038/nm1268

18. Zhong J, Gastaminza P, Cheng G, et al. Robust hepatitis C virus infection in vitro. *Proc Natl Acad Sci U S A*. 2005. doi:10.1073/pnas.0503596102

19. Willingham MC, Pastan IH. Transit of epidermal growth factor through coated pits of the golgi system. *J Cell Biol*. 1982. doi:10.1083/jcb.94.1.207

20. Bananis E, Murray JW, Stockert RJ, Satir P, Wolkoff AW. Microtubule and motor-dependent endocytic vesicle sorting in vitro. *J Cell Biol*. 2000. doi:10.1083/jcb.151.1.179

21. Murray JW, Sarkar S, Wolkoff AW. Single vesicle analysis of endocytic fission on microtubules in vitro. *Traffic*. 2008. doi:10.1111/j.1600-0854.2008.00725.x

22. Mukhopadhyay A, Quiroz JA, Wolkoff AW. Rab1a regulates sorting of early endocytic vesicles. *Am J Physiol Gastrointest Liver Physiol*. 2014;306(5):G412-24. doi:10.1152/ajpgi.00118.2013

23. Farquhar MJ, Hu K, Harris HJ, et al. Hepatitis C Virus Induces CD81 and Claudin-1 Endocytosis. *J Virol*. 2012. doi:10.1128/JVI.06996-11

24. Ploss A, Evans MJ, Gaysinskaya VA, et al. Human occludin is a hepatitis C virus entry factor required for infection of mouse cells. *Nature*. 2009;457(7231). doi:10.1038/nature07684

25. Seeff LB, Lindsay KL, Bacon BR, Kresina TF, Hoofnagle JH. Complementary and alternative medicine in chronic liver disease. *Hepatology*. 2001. doi:10.1053/jhep.2001.27445

26. Lin CW, Tsai FJ, Tsai CH, et al. Anti-SARS coronavirus 3C-like protease effects of Isatis indigotica root and plant-derived phenolic compounds. *Antiviral Res*. 2005;68(1). doi:10.1016/j.antiviral.2005.07.002

27. Zandi K, Teoh BT, Sam SS, Wong PF, Mustafa M, Abubakar S. Antiviral activity of four types of bioflavonoid against dengue virus type-2. *Virol J*. 2011. doi:10.1186/1743-422X-8-560

28. Gutiérrez RMP, Mitchell S, Solis RV. Psidium guajava: A review of its traditional uses, phytochemistry and pharmacology. *J Ethnopharmacol*. 2008. doi:10.1016/j.jep.2008.01.025

29. Khan MR, Kihara M, Omoloso AD. Antimicrobial activity of Michelia champaca. *Fitoterapia*. 2002. doi:10.1016/S0367-326X(02)00248-4

30. Ayyanar M, Subash-Babu P. Syzygium cumini (L.) Skeels: A review of its phytochemical constituents and traditional uses. *Asian Pac J Trop Biomed*. 2012. doi:10.1016/S2221-1691(12)60050-1

31. De Caluwé E, Halamová K, Van Damme P. Tamarindus indica L. – A review of traditional uses, phytochemistry and pharmacology. *Afrika Focus*. 2010;23(1). doi:10.21825/af.v23i1.5039

32. Rost B, Sander C. Conservation and prediction of solvent accessibility in protein families. *Proteins Struct Funct Bioinforma*. 1994;20(3). doi:10.1002/prot.340200303

33. Kitadokoro K, Bordo D, Galli G, et al. CD81 extracellular domain 3D structure: Insight into the tetraspanin superfamily structural motifs. *EMBO J*. 2001;20(1-2). doi:10.1093/emboj/20.1.12

34. Rajesh S, Sridhar P, Tews BA, et al. Structural Basis of Ligand Interactions of the Large Extracellular Domain of Tetraspanin CD81. *J Virol*. 2012;86(18). doi:10.1128/jvi.00559-12

35. Bartosch B, Cosset FL. Studying HCV cell entry with HCV pseudoparticles (HCVpp). *Methods Mol Biol*. 2009;510. doi:10.1007/978-1-59745-394-3_21

36. Pandey MK, Pant AB, Das M. In vitro cytotoxicity of polycyclic aromatic hydrocarbon residues arising through repeated fish fried oil in human hepatoma Hep G2 cell line. *Toxicol Vitr*. 2006;20(3). doi:10.1016/j.tiv.2005.08.005

37. Rost B. Twilight zone of protein sequence alignments. *Protein Eng*. 1999;12(2). doi:10.1093/protein/12.2.85

38. Šali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol*. 1993;234(3). doi:10.1006/jmbi.1993.1626

39. Kong L, Giang E, Nieusma T, et al. Hepatitis C virus E2 envelope glycoprotein core structure. *Science (80-)*. 2013. doi:10.1126/science.1243876

40. Barone D, Balasco N, Autiero I, Vitagliano L. The dynamic properties of the Hepatitis C Virus E2 envelope protein unraveled by molecular dynamics. *J Biomol Struct Dyn*. 2017;35(4). doi:10.1080/07391102.2016.1162198

8. Tables and Figures

Table 1. Data-bases, online tools and softwares used in the study

Online tool / software used	Database used	Purpose	Reference
	NCBI	Retrieval of protein sequences	http://www.ncbi.nlm.nih.gov/
	UniProt	Retrieval of protein sequences	http://www.uniprot.org
BLAST(tblastn)		Obtain corresponding nucleotide sequences from protein sequences by sequence similarity search.	https://blast.ncbi.nlm.nih.gov/Blast.cgi
Gene Infinity (2016 Gene InfinityLLC)		Quantification of codon usage and their frequency	http://www.geneinfinity.org/sms/sms_codonusage.html
mega7 software (version MEGA 7.0.26)		Analysis and quantification of nucleotide sequences. Frequency of different nitrogenous bases present in E2 like Adenine, Thymine, Guanine and Cytosine at their 1^{st}, 2^{nd} and 3^{rd} position can be determined.	[1]
XtalPred server		Characterisation of HCV-E2 and receptor protein CD81 protein sequences by prediction of protein crystallizability.	[2]
Expasy-ProtParam tool		Characterisation of HCV-E2 and receptor protein CD81 protein sequences to analyse molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY)	[3]
Discovery Studio (DS) 2.5		Structural visualisation and analysis of physiological P^H and ionic strength; loop refinement of E2 and CD81 protein structure. Molecular docking between CD81 and E2.	[4]
CASP performers MetaDisorder		Prediction of intrinsic disordered regions of E2 and CD81 structures.	[5]
DISOPRED3		Prediction of intrinsic disordered regions of E2 and CD81 structures.	[6]

TMHMM 2.0		Prediction of transmembrane region of E2 proteins for HCV genotype 1 and 3.	[7]
Phobius		Prediction of transmembrane regions	[8]
	Protein Data Bank (PDB)	Obtaining theoretical protein structure of CD81 (PDB ID: 2AVZ) and truncated E2 structures (PDB ID: 4MWF, chain C and D and 4WEB, chain E) for the validation of de-novo E2-modelles.	[9]
I-TASSER		de-novo modelling of HCV-E2 for the genotype 1 and 3	[10]
Phyre2		de-novo modelling of HCV-E2 for the genotype 1 and 3	[11]
RaptorX		de-novo modelling of HCV-E2 for the genotype 1 and 3	[12]
Verify3D		All the structures retrieved from different modelling servers, validated for stereochemistry, three dimensional structural propensity, non-bonded interactions and to check atomic volumes.	[13]
RAMPAGE		All the structures retrieved from different modelling servers, validated for stereochemistry, three dimensional structural propensity, non-bonded interactions and to check atomic volumes.	[14]
WHAT_CHECK		All the structures retrieved from different modelling servers, validated for stereochemistry, three dimensional structural propensity, non-bonded interactions and to check atomic volumes.	[15]
ERRAT		All the structures retrieved from different modelling servers, validated for stereochemistry, three dimensional structural propensity, non-bonded interactions and to check atomic volumes.	[16]
PROtein Volume Evalution (PROVE)		All the structures retrieved from different modelling servers, validated for stereochemistry, three dimensional structural propensity, non-bonded interactions and to check atomic volumes.	[17]
meta-PPISP		prediction of CD81 binding sites in E2 of genotype 1 and 3	[18]

BSpred		prediction of CD81 binding sites in E2 of genotype 1 and 3	[19]
ConSurf 3.0		prediction of CD81 binding sites in E2 of genotype 1 and 3	[20]

Footnote:

1. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol*. 2016. doi:10.1093/molbev/msw054
2. Slabinski L, Jaroszewski L, Rychlewski L, Wilson IA, Lesley SA, Godzik A. XtalPred: A web server for prediction of protein crystallizability. *Bioinformatics*. 2007;23(24). doi:10.1093/bioinformatics/btm477
3. Gasteiger E, Hoogland C, Gattiker A, et al. Protein Identification and Analysis Tools on the ExPASy Server. In: *The Proteomics Protocols Handbook*. ; 2005. doi:10.1385/1-59259-890-0:571
4. D Studio, . *Discovery Studio Life Science Modeling and Simulations*.; 2008.
5. Kozlowski LP, Bujnicki JM. MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins. *BMC Bioinformatics*. 2012;13(1). doi:10.1186/1471-2105-13-111
6. Jones DT, Cozzetto D. DISOPRED3: Precise disordered region predictions with annotated protein-binding activity. *Bioinformatics*. 2015;31(6). doi:10.1093/bioinformatics/btu744
7. Sönnichsen B, De Renzis S, Nielsen E, Rietdorf J, Zerial M. Distinct membrane domains on endosomes in the recycling pathway visualized by multicolor imaging of Rab4, Rab5, and Rab11. *J Cell Biol*. 2000;149(4). doi:10.1083/jcb.149.4.901
8. Käll L, Krogh A, Sonnhammer ELL. Advantages of combined transmembrane topology and signal peptide prediction-the Phobius web server. *Nucleic Acids Res*. 2007;35(SUPPL.2). doi:10.1093/nar/gkm256
9. Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. *Nucleic Acids Res*. 2000;28(1). doi:10.1093/nar/28.1.235
10. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER suite: Protein structure and function prediction. *Nat Methods*. 2014;12(1). doi:10.1038/nmeth.3213
11. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*. 2015;10(6). doi:10.1038/nprot.2015.053
12. Källberg M, Wang H, Wang S, et al. Template-based protein structure modeling using the RaptorX web server. *Nat Protoc*. 2012;7(8). doi:10.1038/nprot.2012.085
13. Lüthy R, Bowie JU, Eisenberg D. Assessment of protein models with three-dimensional profiles. *Nature*. 1992;356(6364). doi:10.1038/356083a0
14. Lovell SC, Davis IW, Arendall WB, et al. Structure validation by Cα geometry: φ,ψ and Cβ deviation. *Proteins Struct Funct Genet*. 2003;50(3). doi:10.1002/prot.10286
15. Hooft RWW, Vriend G, Sander C, Abola EE. Errors in protein structures [3]. *Nature*. 1996;381(6580). doi:10.1038/381272a0
16. Colovos C, Yeates TO. Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Sci*. 1993;2(9). doi:10.1002/pro.5560020916
17. Pontius J, Richelle J, Wodak SJ. Deviations from standard atomic volumes as a quality measure for protein crystal structures. *J Mol Biol*. 1996;264(1). doi:10.1006/jmbi.1996.0628
18. Qin S, Zhou HX. Meta-PPISP: A meta web server for protein-protein interaction site prediction. *Bioinformatics*. 2007;23(24). doi:10.1093/bioinformatics/btm434
19. Mukherjee S, Zhang Y. Protein-protein complex structure predictions by multimeric threading and template recombination. *Structure*. 2011;19(7). doi:10.1016/j.str.2011.04.006
20. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. ConSurf 2010: Calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res*. 2010;38(SUPPL. 2). doi:10.1093/nar/gkq399

Table 2. Nucleotide sequences of HCV whole genome, RdRp, E1 and E2 used in the Study.

S. no	Genotype	Strain/ isolate	NCBI Accession number	Length of Whole Genome (nucleotide)	Position of E1 (nucleotide)	Position of E2 (nucleotide)	Position of RdRp (nucleotide)		
1	1a	H77	AF011753.1	1 - 9600.	918 - 1487	1497 - 2528	7608 - 9144		
2	1a	HCV-RMT	AB520610.1	1 - 9598.	918 - 1487	1497 - 2528	7608 - 9144		
3	1a	H77	AF011751.1	1 - 9599.	918 - 1487	1497 - 2528	7608 - 9144		
4	1a	H77C	KP098533.1	1 - 9599.	918 - 1487	1497 - 2528	7608 - 9144		
5	1a	H77	AF011752.1	1 - 9599.	918 - 1487	1497 - 2528	7608 - 9144		
6	1a	H77	JX472009.1	1 - 9022.	918 - 1487	1497 - 2528	-		
7	1a	H77	JX472010.1	1 - 9022.	918 - 1487	1497 - 2528	-		
8	1a	Bole1a	JQ791196.1	1 - 9607.	414 - 1490	1497 - 2528	7608 - 9144		
9	1a	HC-TN	EF621489.1	1 - 9599.	918 - 1487	1497 - 2528	7608 - 9144		
10	1a	H77	AF009606.1	1 - 9646.	918 - 1487	1497 - 2528	7608 - 9144		
S. no	**Genotype**	**Strain/ isolate**	**NCBI Accession number**	**Length of Whole Genome (nucleotide)**	**Position of E1 (nucleotide)**	**Position of E2 (nucleotide)**	**Position of RdRp (nucleotide)**		
11	1b	US/BID-V121/1992	EU155324.2	1 - 9329.	866 - 1435	1445 - 2476	7553 - 9088		
12	1b	US/BID-V131/1990	EU155331.2	1 - 9329.	866 - 1435	1445 - 2476	7553 - 9088		
13	1b	US/BID-V141/1990	EU155337.2	1 - 9329.	866 - 1435	1445 - 2476	7553 - 9088		
14	1b	CH/BID-V294/2002	EU155366.2	1 - 9329.	866 - 1435	1445 - 2476	7553 - 9088		
15	1b	US/BID-V132/1996	EU155332.2	1 - 9329.	866 - 1435	1445 - 2476	7553 - 9088		
16	1b	US/BID-V126/1991	EU234061.2	1 - 9329.	866 - 1435	1445 - 2476	7553 - 9088		

17	1b	US/BID-V349/2002	EU256100.1	1 - 9329.	866 - 1435	1445 - 2476	7553 - 9088		

S. no	**Genotype**	**Strain/ isolate**	**NCBI Accession number**	**Position of E2 (nucleotide)**	**Strain/ isolate**	**NCBI Accession number**	**Length of Whole Genome (nucleotide)**	**Position of E1 (nucleotide)**	**Position of RdRp (nucleotide)**
18	3a	NZL1	D17763.1	1156 - 2205	NZL1	D17763.1	1 - 9456.	916 - 1485	7636 - 9171
19	3a	3a	LC217312.1	1495 – 2544	ZS633	KC844041.1	1 - 9434.	916 - 1485	7636 - 9174
20	3a	Pk-E2	GU736410.1	1495 – 2544	ILBSRAS_JK	KF035124.1	1 - 9426.	916 - 1485	7642 - 9177
21	3a	3a-1007	GQ379231.1	1495 – 2544	ILBSRAS_BD	KF035125.1	1 - 9426.	916 - 1485	7639 - 9174
22	3a	It-13	GQ356461.1	1495 – 2544	ILBSRAS_LD	KF035126.1	1 - 9426.	916 - 1485	7642 - 9177
23	3a	It-16	GQ356661.1	1495 - 2544	ILBSRAS_UG	KF035127.1	1 - 9426.	916 - 1485	7642 - 9177
24	3a	It-14	GQ356564.1	1495 – 2544					
25	3a	s48_v01_d84_c1	GU055318.1	1495 – 2544					

S. no	**Genotype**	**Strain/ isolate**	**NCBI Accession number**	**Length of Whole Genome (nucleotide)**	**Position of E1 (nucleotide)**	**Position of E2 (nucleotide)**	**Position of RdRp (nucleotide)**		
26	3b	ZS650	KC844044.1	1 - 9434.	916 - 1485	1498 – 2547	7640 - 9174		
27	3b	QC34	KJ470616.1	1 - 9458.	916 - 1485	1498 – 2547	7636 - 9171		
28	3b	TrKj	D49374.1	1 - 9444.	916 - 1488	1498 – 2547	7642 - 9177		
29	3b	SH37	JQ065709.1	1 - 9383.	916 - 1488	1498 - 2547	7614 - 9149		

Table 3. Nucleotide composition analysis (Mean±SD) of HCV glycoprotein E2 from genotypes 1 and 3.

Genotype	%A	%C	%T	%G	SD within a genotype
1a	20.2±0.71	30.5±0.39[a, b, c]	21.8±0.84	27.5±0.18	4.28
1b	20.3±0.59	30.0±0.83[a, b, c]	21.5±0.53	28.2±0.37	4.28
3a	21.3±0.31	27.9±0.68[a, b, d]	24. 6±0.85	26.3±0..54	2.55
3b	21.8±1.01	27.9±0.65[a, b, d]	24.2±1.31	26.1±1.42	2.52

a: Significant difference between Cytosine and other nucleotides within a strain, p<0.01
b: Significant differences between Cytosine content of genotype 1 and 3, p<0.01
c: Insignificant differences between Cytosine content within genotype 1, p>0.01
d: Insignificant differences between Cytosine content within genotype 3, p>0.01
One way analysis of variance (ANOVA) was applied and comparison was made through Bonferroni's test.

Table 4. Percent of mean nucleotide content in whole genome of Hepatitis C virus.

Genotype	Mean %T for all genotype	Mean %C for all genotypes	Mean %A for all genotypes	Mean %G for all genotype	Standard deviation of the nucleotide variation for each genotype.
1a	21.7±0.3	30.2±0.12[a,b]	19.7±0.1	28.4±0.18	4.44
1b	20.93±0.75	30.3±0.05[a,c]	20.1±0.13	28.6±0.13	4.59
3a	23.18±0.18	28.2±0.15[a, d]	21.0±0.13	27.5±0.21	3.3
3b	22.18±0.14	28.7±0.1[a,e]	21.4±0.22	27.7±0.22	3.33

a: Significant differences amongst other nucleotides in the same genome (p<0.05).
b: significant difference from Cytosine content of genotype 3a and 3b (p<0.05), insignificant from 1b.
c: significant difference from Cytosine content of genotype 3a and 3b (p<0.05), insignificant from 1a.
d: significant difference from Cytosine content of genotype 1a and 1b (p<0.05), insignificant from 3b.
e: significant difference from Cytosine content of genotype 1a and 1b (p<0.05), insignificant from 3a.
One way analysis of variance (ANOVA) was applied and comparison was made through Bonferroni's test.

Table 5. Percent of mean nucleotide content in E1.

Genotype	Mean %T for all genotype	Mean %C for all genotypes	Mean %A for all genotypes	Mean %G for all genotype	Standard deviation of the nucleotide variation for each genotype.
1a	23.502±1.01	29.5±0.52[a,b]	17.5±0.83	29.5±0.52	5.09
1b	23.659±0.62	29.5±0.52[a,b]	18.1±0.34	28.8±0.33	4.67
3a	25.124±0.50	27.0±0.75[a,c]	19.8±0.45	28.1±0.45	3.28
3b	22.656± 0.472	28.3±0.24[a, d]	21.1±0.9	28.0±0.83	3.34

a: Significant differences amongst other nucleotides in the same genome (p<0.05).
b: significant difference from Cytosine content of genotype 3a and 3b (p<0.05), insignificant from 1b.
c: significant difference from Cytosine content of genotype 3a and 3b (p<0.05), insignificant from 1a.
d: significant difference from Cytosine content of genotype 1a and 1b (p<0.05), insignificant from 3b.
e: significant difference from Cytosine content of genotype 1a and 1b (p<0.05), insignificant from 3a.
One way analysis of variance (ANOVA) was applied and comparison was made through Bonferroni's test.

Table 6. Percent of mean nucleotide content in RdRp.

Genotype	Mean %T for all genotype	Mean %C for all genotypes	Mean %A for all genotypes	Mean %G for all genotype	Standard deviation of the nucleotide variation for each genotype.
1a	19.9±0.118	29.9±0.092[a,b]	23.9±0.32	26.3±0.3	3.69
1b	20.6±0.161	29.5±0.20[a,c]	23.9±0.18	25.9±0.15	3.28
3a	22.0±0.266	28.6±0.08[a, d]	22.7±0.27	26.7±0.3	2.81
3b	21.9±0.125	28.6±0.31[a,e]	23.4±0.43	26.1±0.25	2.64

a: Significant differences amongst other nucleotides in the same genome (p<0.05).
b: significant difference from Cytosine content of genotype 3a and 3b (p<0.05), insignificant from 1b.
c: significant difference from Cytosine content of genotype 3a and 3b (p<0.05), insignificant from 1a.
d: significant difference from Cytosine content of genotype 1a and 1b (p<0.05), insignificant from 3b.
e: significant difference from Cytosine content of genotype 1a and 1b (p<0.05), insignificant from 3a.
One way analysis of variance (ANOVA) was applied and comparison was made through Bonferroni's test.

Table 7. Nucleotide content at different codon positions in HCV-E2.

	Percentage occurrence of nucleotide at indicated positions in codons											
		1st position (%)				2nd position (%)				3rd position (%)			
Genotype	Number of sequences analysed	T	C	A	G	T	C	A	G	T	C	A	G
1a Average	10	22.8± 1.48	22.2± 0.46*	26.1± 1.48	28.8± 0.2	24.4± 1.42	25.3± 0.43*	23.2± 1.06	26.9± 0.71	18± 0.88	44.1± 0.94[a, b]	11.1± 0.63	26.7± 0.82
SD		2.88				1.65				12.55			
1b Average	7	22.8± 0.90	22.6± 0.72*	26.6± 0.85	27.9± 0.49	24± 0.00	27.1± 0.67 [a]	24± 0.33	24.7± 0.47	18± 1.62	40.3± 2.31 [a, b]	10.2± 1.31	31.9± 0.91
SD		2.44				1.38				12.11			
3a Average	8	23.1± 1.25	27± 1.05*, b	21.8± 0.68	28.2± 0.84	23.12± 1.24	27± 1.06*	21.8± 0.68	21.8± 0.68	27± 0.93	29.8± 0.92 [a, b]	19± 0.98	24.2± 0.83
SD		2.84				2.84				4.14			
3b Average	4	23.8± 0.96	22.4± 0.51*	25.4± 1.07	28.3± 1.66	23.8± 0.96	22.4± 0.5*.b	25.4± 1.06	28.3± 1.66	20.2±3.0	36.3± 1.98 [a, b]	17± 1.25	26.7± 2.0
SD		2.51				2.5				7.86			

***a**: Significant difference between Cytosine and other nucleotides within a strain, p<0.01*

***b**: Significant differences between Cytosine content of other genotypes, p<0.01*

***not significant against all other nucleotides within a genotype. Analysis was done via* one way ANOVA followed by Bonferonni post hoc analysis.**

Table 8. Nucleotide content at different codon positions in HCV whole genome.

Percentage occurrence of nucleotide at indicated positions in codons													
		1st position (%)				2nd position (%)				3rd position (%)			
Genotype	Number of sequences analysed	T	C	A	G	T	C	A	G	T	C	A	G
1a Average	10	27.2±0.32	28.5±0.1	21.8±0.22	22.6±0.14	19.0±057	38.1±0.21[a,b]	13.5±0.12	29.4±0.3	18.9±0.42	24.0±0.10	23.9±0.14	33.2±0.18
SD		2.89				9.6				5.24			
1b Average	7	17.1±0.0	37.9±0.18	14.3±0.3	30.7±0.32	18.9±0	23.8±0.09[*, b]	23.7±0.17	33.6±0.17	26.7±0.0	29.3±0.13[a,b]	22.5±0.13	21.5±0.13
SD		9.89				5.41				3.25			
3a Average	6	19.8±0.40	23.8±0.18	23.5±0.23	33.0±0.2	26.8±0.40	28.7±0.11[a, *]	22.6±0.33	22.3±0.16	23.2±0.41	32.3±0.42[a,b]	17.3±0.31	27.3±0.41
SD		4.98				2.83				5.64			
3b Average	4	20.0±0	23.7±0.14	23.3±0.18	33.2±0.3	27.0±0	28.4±0.10[a, *]	22.6±0.33	22.2±0.13	19.9±0.0	33.9±0.40[a,b]	18.4±0.29	27.8±0.34
SD		5.06				2.81				6.46			

a: Significant difference between Cytosine and other nucleotides within a strain, p<0.01

b: Significant differences between Cytosine content of other genotypes, p<0.01

****not significant against all other nucleotides within a genotype. Analysis was done via* one way ANOVA followed by Bonferonni post hoc analysis**

Table 9. Nucleotide content at different codon positions in HCV-E1 genome. Percentage occurrence of nucleotide at indicated positions in codons

		1st position (%)				**2nd position (%)**				**3rd position (%)**			
Genotype	Number of sequences analysed	T	C	A	G	T	C	A	G	T	C	A	G
1a Average	10	17.7± 2.11	24.7±0. 22	23.7±1. 84	33.7±0. 78	31.6±0.4 7	27.0±0. 51	19.8±0. 47	21.7±0. 38	20.9±1.7 9	36.9±1. 59[a,*]	8.93±1. 0	33.05±1 .0
SD		5.93				4.69				11.2			
1b Average	7	21±0. 30	20.9±0. 29	21.4±0. 54	36.5±0. 65	32.6±0.5 3	27.5±0. 41	20.5±0. 0	19.5±0. 0	17.14±1. 46	39.93±1 .24[a,*]	12.39±0 .27	30.46±0 .57
SD		6.73				5.45				11.12			
3a Average	6	21±0. 55	19.9±0. 64	24.2±1. 06	35.4±0. 72	31.5±0.5 5	26.5±0. 41	21.0±0. 42	21.3±0. 44	23.7±1.3 7	34.8±1. 67[a, **]	14.2±1. 17	27.47±1 .45
SD		6.42				4.43				7.67			
3b Average	4	23±0. 82	19.3±0. 29	24.2±1. 16	33.5±1. 49	32.3±0.5	26.7±0. 66	21.4±1. 6	19.9±1. 79	13±0.81	38.8±0. 42[a,***]	17.7±2. 73	30.7±2. 46
SD		5.48				5.14				10.72			

a: Significant difference between Cytosine and other nucleotides within a strain, p<0.01

b: Significant differences between Cytosine content of other genotypes, p<0.01

****not significant against all other nucleotides within a genotype.***

***** not significant against genotype 1a, significant against other genotypes***

****** not significant against genotype 1 and b, significant against genotype 3a. Analysis was done via* one way ANOVA followed by Bonferonni post hoc analysis.**

Table 10. Nucleotide content at different codon positions in HCV-RdRp genome. Percentage occurrence of nucleotide at indicated positions in codons

		1st position (%)				**2nd position (%)**				**3rd position (%)**			
Genotype	Number of sequences analysed	T	C	A	G	T	C	A	G	T	C	A	G
1a Average	8	18.0± 0.0	23.6± 0.07	26.0± 0.1	32.4± 0.28	25.0± 0.0	26.9± 0.21	26.6± 0.29	21.2± 0.15	16.6± 0.52	39.2± 0.26[a,b]	19.1± 0.62	25.3± 0.6
SD		5.27				2.3				8.91			
1b Average	7	19.1± 0.38	23.9± 0.34	26.1± 0.22	30.7± 0.31	25.0± 0.0	27.5± 0.16	27.7± 0.36	19.9± 0.3	17.5± 0.53	37.1± 0.48[a,b]	17.9± 0.53	27.4± 0.47
SD		4.24				3.34				8.19			
3a Average	6	19.0± 0.00	23.4± 0.18	26.3± 0.44	31.5± 0.15	25.3± 0.52	28.2± 0.24	26.8± 0.22	19.6± 0.21	21.8± 0.63	34.2± 0.39[a,c]	14.9± 0.62	29.0± 0.91
SD		4.63				3.36				7.47			
3b Average	4	20.3± 0.5	22.5± 0.23	26.2± 0.28	31.2± 0.3	25.3± 0.5	27.6± 0.26	27.2± 0.25	19.9± 0.34	20.5± 0.58	35.6± 0.86[a,c]	16.8± 1.24	27.3± 0.91
SD		4.28				3.2				7.43			

a: Significant difference between Cytosine and other nucleotides within a strain, p<0.05

b: Significant differences between Cytosine content of other genotypes, p<0.05

c: Significant differences between Cytosine content of 1a and 1b, p<0.05. ***Analysis was done via*** **one way ANOVA followed by Bonferonni post hoc analysis**

Table 11. Codon usage of HCV-E2 across HCV Genotypes. Values in bold indicate the most abundant codon present in E2, red are abundant codons that are not GC rich.

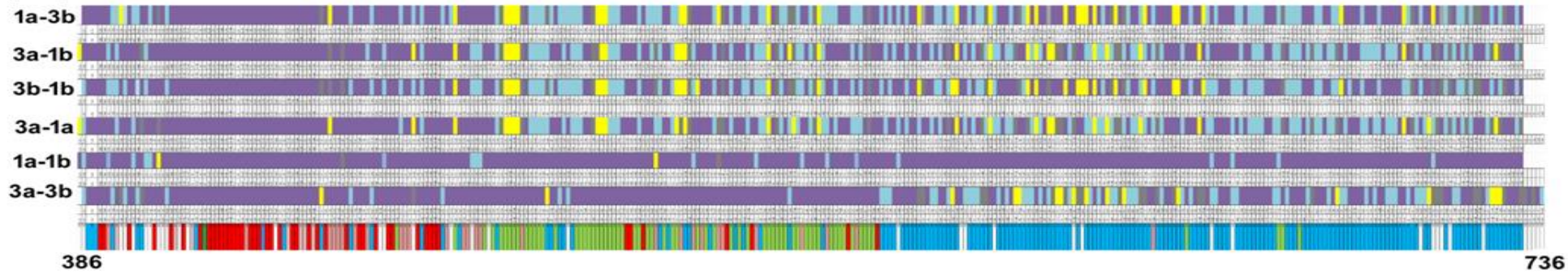
		1a	1b	3a	3b
Amino acid	**Codon**	**Average ± SD**	**Average ± SD**	**Average ± SD**	**Average ± SD**
Ala	GCG	24.9±5.94	**34.9±9.87**	16.8±5.54	17.1±3.76
Ala	GCA	20.2±3.91	18.3±9.27	27.9±7.51	14.6±2.42
Ala	GCT	7.2±3.83	15.1±6.19	**34.8±3.04**	26.4±7.60
Ala	GCC	**47.7±7.07**	31.8±6.74	20.5±6.46	**41.9±6.39**
Cys	TGT	44.9±3.43	31.0±7.07	33.8±3.47	32.1±9.38
Cys	TGC	**55.1±3.43**	**69.0±7.07**	**66.2±3.47**	**67.9±9.38**
Asp	GAT	36.6±6.94	27.5±12.20	**51.7±8.44**	**57.5±6.38**
Asp	GAC	**63.4±6.94**	**72.5±12.20**	48.3±8.44	42.5±6.38
Glu	GAG	46.9±7.11	**96.0±5.03**	**86.9±7.76**	**70.5±2.88**
Glu	GAA	**53.1±7.11**	4.0±5.03	13.1±7.76	29.5±2.88
Phe	TTT	12.9±5.16	24.0±3.53	35.8±8.19	37.8±6.93
Phe	TTC	**87.1±5.16**	**76.0±3.53**	**64.2±8.19**	**62.2±6.93**
Gly	GGG	19.3±2.58	**43.0±1.62**	**31.1±5.21**	**39.0±6.55**
Gly	GGA	22.6±3.40	8.9±1.60	13.8±4.36	18.6±4.93
Gly	GGT	14.7±1.66	19.6±4.95	24.7±5.97	15.7±3.14
Gly	GGC	**43.4±1.33**	28.5±3.88	30.4±4.89	26.8±1.98
His	CAT	16.5±7.85	25.0±13.65	**50.9±11.44**	34.4±3.59
His	CAC	**83.5±7.85**	**75.0±13.65**	49.1±11.44	**65.6±3.59**
Ile	ATA	10.4±8.05	21.2±6.99	32.8±5.47	**47.3±3.45**
Ile	ATT	28.5±6.30	12.1±7.27	7.4±8.62	13.7±10.30
Ile	ATC	**61.1±12.67**	**66.7±6.52**	**59.8±7.76**	39.1±10.52
Lys	AAG	43.3±7.56	**68.1±13.13**	**60.3±5.60**	**72.5±10.02**
Lys	AAA	**56.7±7.56**	31.9±13.13	39.7±5.60	27.5±10.02
Leu	TTG	18.2±1.10	20.6±3.80	15.9±5.08	21.1±7.53
Leu	TTA	2.9±1.81	3.0±1.49	4.1±2.47	3.0±2.17
Leu	CTG	**29.0±4.60**	26.0±10.16	**31.2±7.37**	19.8±3.59
Leu	CTA	0.3±1.05	7.5±4.18	8.2±5.26	14.7±6.45
Leu	CTT	25.2±2.47	16.4±4.10	17.0±3.58	16.4±2.96
Leu	CTC	24.3±4.05	**26.4±6.39**	23.6±3.84	**24.9±2.64**
Met	ATG	100.0±0.00	100.0±0.00	100.0±0.00	100.0±0.00

Asn	AAT	27.3±5.15	25.6±5.85	36.8±8.31	30.1±3.87
Asn	AAC	**72.7±5.15**	**74.4±5.85**	**63.2±8.31**	**69.9±3.87**
Pro	CCG	22.6±2.08	27.9±4.11	8.9±3.92	14.0±2.10
Pro	CCA	15.8±2.63	15.1±4.32	20.1±1.77	29.4±3.70
Pro	CCT	26.2±3.32	24.0±4.69	32.4±5.82	14.9±3.23
Pro	CCC	**35.5±4.11**	**33.0±7.85**	**38.6±7.04**	**41.7±4.50**
Gln	CAG	**82.0±8.44**	**78.6±10.39**	41.8±7.58	38.1±20.14
Gln	CAA	18.0±8.44	21.4±10.39	**58.2±7.58**	**61.9±20.14**
Arg	AGG	**50.0±1.91**	**39.1±5.27**	**27.8±4.93**	**25.9±4.35**
Arg	AGA	4.4±2.35	9.5±4.07	16.9±6.54	16.3±7.24
Arg	CGG	12.4±3.92	13.4±3.78	23.4±3.69	13.8±6.21
Arg	CGA	5.6±0.20	11.1±4.49	4.7±3.59	11.5±8.68
Arg	CGT	2.9±5.00	9.6±4.31	15.2±4.16	7.2±3.90
Arg	CGC	24.5±6.42	17.3±6.46	12.0±6.05	25.3±3.31
Ser	AGT	10.9±5.33	6.5±5.62	**19.8±8.12**	10.3±6.90
Ser	AGC	**34.3±4.97**	**28.0±7.98**	15.9±4.51	16.8±10.23
Ser	TCG	4.1±1.77	17.1±6.53	11.1±3.81	11.9±6.12
Ser	TCA	13.2±1.90	11.0±5.51	18.3±4.72	**25.5±7.36**
Ser	TCT	12.3±2.80	7.5±2.63	17.5±4.60	14.7±3.37
Ser	TCC	25.2±5.45	**29.9±4.52**	17.5±3.49	20.8±7.79
Thr	ACG	22.2±2.92	21.2±6.26	15.6±2.90	23.7±7.32
Thr	ACA	11.4±3.45	18.0±3.18	26.3±4.81	22.1±6.97
Thr	ACT	12.2±3.10	21.5±4.17	24.8±6.53	20.4±7.95
Thr	ACC	**54.1±2.37**	**39.3±7.48**	**33.3±5.95**	**33.9±7.51**
Val	GTG	**46.8±3.32**	**48.6±4.43**	**38.9±3.92**	**49.8±6.05**
Val	GTA	8.0±2.87	9.9±3.84	7.6±3.62	12.6±4.53
Val	GTT	6.8±2.22	18.5±4.61	15.6±7.48	8.5±5.61
Val	GTC	38.4±1.67	22.9±5.17	37.9±8.23	29.2±6.00
Trp	TGG	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**
Tyr	TAT	38.3±16.12	24.1±10.90	34.9±7.36	36.1±19.42
Tyr	TAC	**61.7±16.12**	**75.9±10.90**	**65.1±7.36**	**63.9±19.42**

Table 12. Accession numbers of the representative strains and position of E2 used for modelling and other analysis.

Genotype	Strain	NCBI Nucleotide Accession No.	NCBI Protein Accession No.	Position of E2 (aa)
1a	H77	AF011753.1	AAB67038.1	386 - 729
1b	CH/BID-V294/2002	EU155366.2	ABV46290.1	386 - 729
3a	NZL1	D17763.1	YP_001469631.1	386 - 735
3b	ZS650	KC844044.1	AHH29579.1	387 - 736

(A)



(B)

Colour Code	Change Indicated	Percentage
	Same amino acid in 1a and 1b but different in 3a and 3b	47% (165)
	Same amino acid in both genotypes 1 and 3	13% (45)
	Same amino acids in 3a and 3b but different in 1a and 1b	10% (36)
	Same amino acids in 3a and 3b, 1a and 1b but genotype1 different from 3	19% (66)
	Uncategorized	11% (38)
	TOTAL LENGTH	**350**

(C)

Colour Code	Amino acid change	3A – 3B	1A – 1B	1A – 3A	3B – 1B	3A – 1B	1A – 3B
	Polar vs non polar	18% (64)	5% (19)	29% (100)	29% (100)	29% (100)	29% (100)
	Aromatic - aliphatic	8% (28)	1% (2)	9% (32)	9% (30)	10% (33)	8% (27)
	Polar, aromatic – non polar aliphatic	5% (17)	1% (2)	8% (28)	9% (31)	8% (27)	10% (33)
	Uncategorized	68% (239)	93% (320)	54% (183)	53% (181)	53% (183)	53% (182)
	Total length	350	344	344	343	344	343
	Acidic-Basic	0.02% (7)	0% (1)	0.01% (3)	0.01% (5)	0.01% (4)	0.02% (6)
	Acidic/Basic-uncharged polar/	0.15% (54)	0.03%	0.21%	0.22%	0.23% (79)	0.2% (68)

Fig1. Amino acids changes between genotype 1 and 3 of HCV-E2. (A) Pictorial representation to indicate the changes in E2 between strains of genotypes 1 and 3. Color coded changes are quantified in (B), where percentage of amino acids changes scored between the four strains are also depicted (C) Percentage of amino acids changes that were considered between the two genotypes.

Table 13. Changes of nucleotide position for each HCV-E2 codon. Percentage has been represented using the longest E2 sequence (350 aa).

Changes	Only 1st nt	Only 2nd nt	Only 3rd nt	All 3 nt	Both 1 and 2nt	Both 1 and 3nt	Both 2 and 3nt	None
1a vs 1b	15 or 2%	7 or 2%	124 or 12%	11 or 1%	9 or 1%	25 or 2%	8 or1%	144 or 14%
3a vs 3b	18 or 2%	21or 2%	70 or 7%	62 or 6%	24 or 2%	45 or 4%	33 or 3%	76 or 7%
3b vs 1a	27or 2%	17 or 2%	37 or 4%	104 or 10%	46 or 4%	41 or 4%	38 or 4%	31 or 3%
1b vs 3a	18 or 2%	17 or 2%	37 or 4%	109 or 10%	35 or 3%	56 or 5%	31 or 3%	37 or 4%
3b vs 1b	20 or 2%	12 or 1%	36 or 3%	113 11%	43 or 4%	45 or 4%	34 or 3%	31 or 3%
3a vs 1a	19 or 2%	10 or 1%	41 or 4%	111 or 11%	38 or 4%	48 or 5%	36 or 3%	34 or 3%

Footnote: The exact number of changed E2-nucleotides along with percentage is shown.

Table 14: Validation of E2 models by different validation tools

Server	**1a_E2**				**1b_E2**				**3a_E2**				**3b_E2**				**CD81**							
RAMPAGE	**Phyre^2**	**Phyre^2$_{EM}$**	**I-TASSER**	**I-TASSER$_{EM}$**	**Phyre^2**	**Phyre^2$_{EM}$**	**I-TASSER**	**I-TASSER$_{EM}$**	**Phyre^2**	**Phyre^2$_{EM}$**	**I-TASSER**	**I-TASSER$_{EM}$**	**Phyre^2**	**Phyre^2$_{EM}$**	**I-TASSER**	**I-TASSER$_{EM}$**	**I-TASSER**	**I-TASSER$_{EM}$**	**Phyre^2**	**Phyre^2$_{EM}$**	**Raptor X**	**Raptor X$_{EM}$**	**PDB ID 2AVZ**	**PDB ID 2AVZ$_{EM}$**
% of Residues in favoured region (≈98%)	72.5	76.9	64.2	76	76	77.1	56.9	57.8	70.6	74.1	70.6	71.5	80.7	81.6	70	71.8	80.7	82	92	91.9	94	94.8	92.3	92.3
% of Residues in allowed region (≈2%)	13.9	15.6	24.3	16.7	12	12	29	27.9	14.4	11	23.1	23.3	13.3	13	22.5	22.5	10.3	9.4	4.6	5.8	5.1	4.7	5.1	4.7
% of Residues in outlier region	13.6	7.5	11.4	7.3	12	10.9	14.1	14.4	15	15	6.3	5.2	6.1	5.5	7.5	5.8	9	8.6	3.4	2.3	0.9	0.4	2.6	3
Overall G-factor (>-0.5)	-2.78	-0.34	-0.83	-0.54	-2.69	-0.8	-1.03	-0.95	-2.76	-1.11	-0.7	-0.65	-0.67	-0.67	-0.63	-0.58	-0.59	-0.46	-0.4	-0.09	-0.06	0.03	0	-0.03

PROVE																							
Z-Score RMS (<1)	2.55	1.51	1.66	1.48	2.904	1.893	-	-	-	1.835	-	-	1.969	1.73	1.724	1.709	1.507	not found	2.277	1.556	1.38	not found	1.44	1.467
ERRAT																								
Overall quality factor	-	81.74	-	69.35	-	52.961	51.057	51.807	-	37.805	66.866	65.563	-	41.765	81.009	79.822	85.903	85.088	71.25	100	89.055	96.392	93.392	95.595
VERIFY3D																								
% of the residues with averaged 3D-1D score ≥0.2 (≈65%)	71.35	74.21	90.7	91.57	64.83	65.41	77.33	80.23	56.86	60.86	89.71	90.29	61.71	63.71	82	79.71	45.34	62.71	35.96	56.18	22.03	22.88	52.97	50.85
WHATCHECK																								
Structure Z-scores(>0):																								
2nd generation packing quality	-4.83	-4.44	-3.86	-3.85	-4.535	-4.118	-3.808	-3.734	-4.834	-4.443	-3.858	-3.85	-4.502	-4.341	-4.065	-3.993	-0.536	-0.323	-2.888	-2.857	-2.051	-1.546	-0.662	-0.688
Ramachandran plot	-6.19	-6.	-6.	-6.	-5.	-5.	-7.	-7.	-6.	-6.	-6.	-6.	-4.	-4.	-6.	-6.	-4.	-4.	-5.	-4.	1.03	0.54	-0.	-0.

appearance		05	38	19	659	188	996	831	195	052	383	194	339	452	157	08	297	252	114	777	6	8	072	274
chi-1/chi-2 rotamer normality	-3.01	-4.1	-4.85	-4.36	-2.564	-3.843	-5.275	-5.299	-3.011	-4.101	-4.853	-4.364	-3.288	-4.537	-4.023	-3.993	-4.548	-4.243	4.518	-1.749	-2.544	-2.554	-4.725	-4.975
Backbone conformation	-4.28	-4.69	-5.99	-5.81	-3.701	-3.892	-8.403	-8.271	-4.28	-4.689	-5.994	-5.81	-6.453	-6.703	-5.977	-5.926	-3.57	-3.361	-3.37	-3.665	-1.177	-1.128	-0.977	-1.043

NOTE: Values of each parameter of the servers were given within bracket in the first column. Server outputs were named after the modelling servers. Energy minimized structures were subscripted as EM after the server name. The chosen models are shaded.

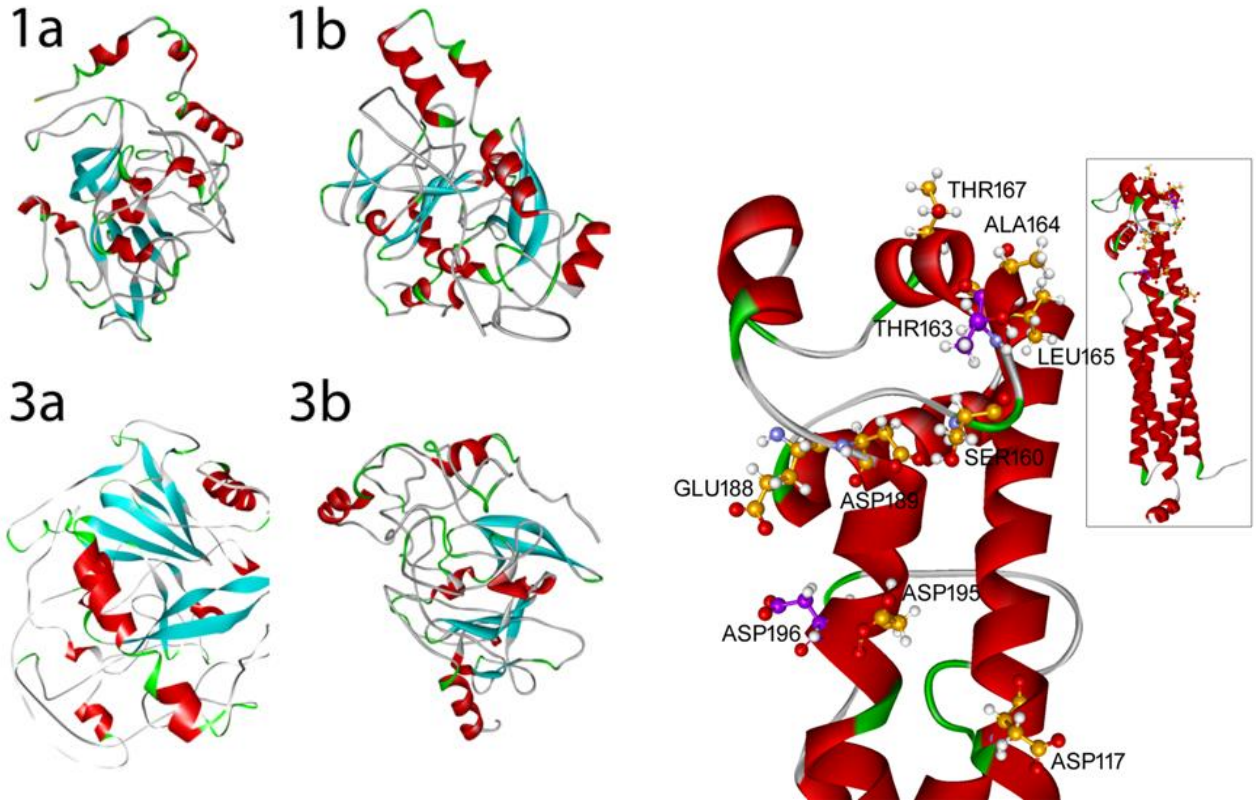


Fig 2. Predicted models of E2 from genotype 1a, 1b, 3a and 3b as indicted above. Figure to the right indicates PDB structure of CD81 (2AVZ).

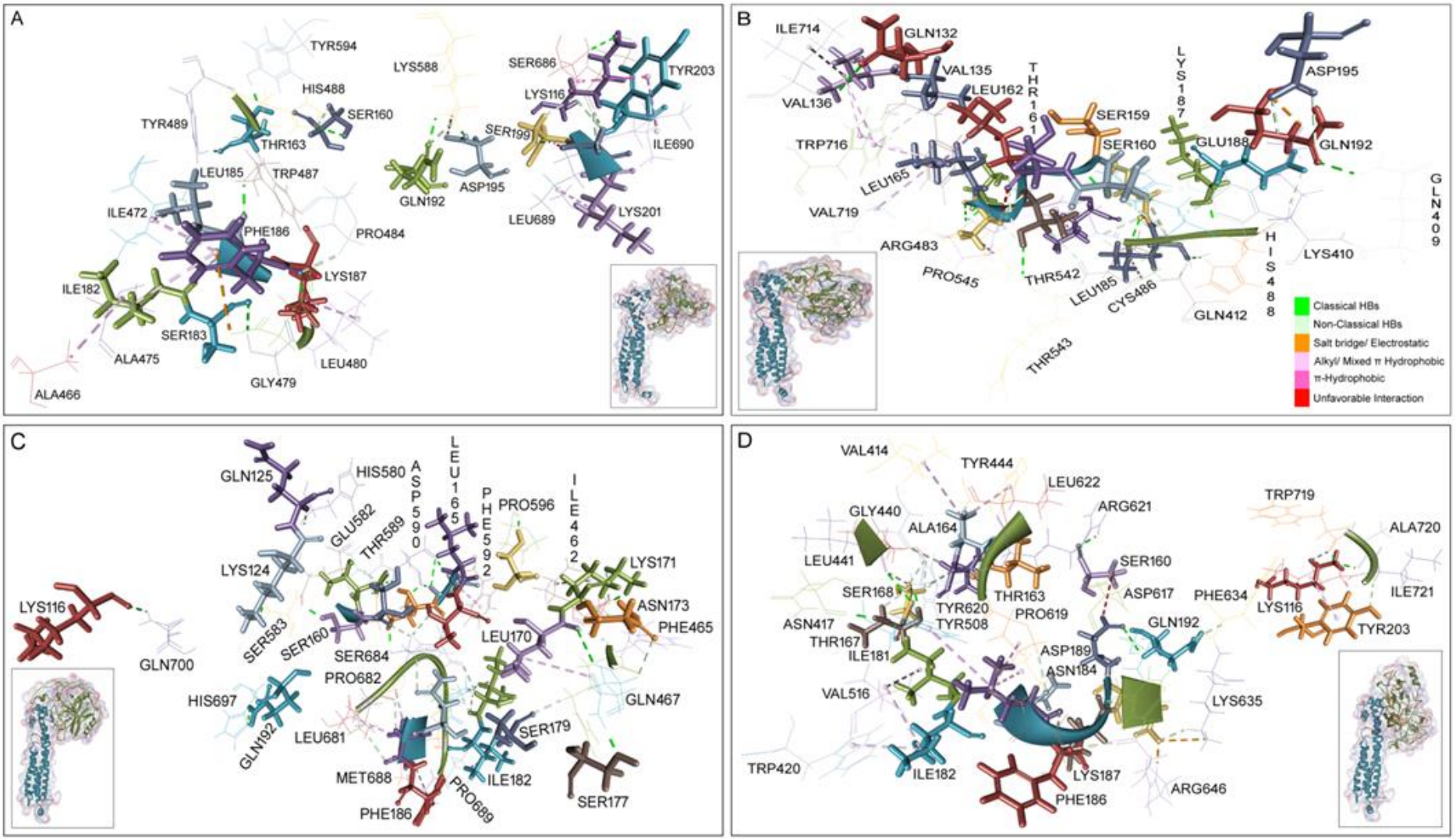


Fig3. Interaction profile of CD81 and HCV-E2. CD81 interaction with E2 is depicted as a 2D snapshot of a randomly selected orientation of genotype 1a (A), genotype 1b (B), genotype 3a (C) and genotype 3b (D).

Table 15. Relative Solvent Accessibility of HCV-E2 residues of genotypes 1 and 3 known to be antigenic determinants. Amino acids are represented by their standard single letter code.

Position	1a		1b		3a		3b	
412	Q	exposed	Q	exposed	Q	exposed	H	exposed
413	L	exposed	L	buried	L	exposed	L	buried
414	I	exposed	I	buried	V	buried	V	buried
415	N	exposed	N	buried	N	exposed	N	buried
416	T	exposed	T	buried	T	exposed	T	buried
417	N	exposed	N	exposed	N	exposed	N	exposed
418	G	exposed	G	buried	G	exposed	G	exposed
419	S	exposed	S	buried	S	buried	S	exposed
420	W	exposed	W	exposed	W	buried	W	exposed
421	H	exposed	H	exposed	H	exposed	H	exposed
422	I	exposed	I	exposed	I	exposed	I	exposed
423	N	exposed	N	exposed	N	exposed	N	exposed
427	L	exposed	L	exposed	L	exposed	L	exposed
428	N	exposed	N	exposed	N	exposed	S	exposed
429	C	exposed	C	exposed	C	exposed	C	exposed
430	N	exposed	N	exposed	N	exposed	N	exposed
431	E	exposed	D	exposed	E	exposed	D	exposed
432	S	exposed	S	exposed	S	exposed	S	exposed
433	L	exposed	L	exposed	I	exposed	L	exposed
434	N	exposed	Q	exposed	N	exposed	N	exposed
435	T	exposed	T	buried	T	exposed	T	exposed
436	G	exposed	G	buried	G	exposed	G	exposed
437	W	buried	F	buried	F	buried	F	buried
438	L	exposed	I	exposed	I	buried	I	exposed
439	A	exposed	A	buried	A	exposed	A	exposed
440	G	buried	A	exposed	G	buried	G	buried
441	L	exposed	L	exposed	L	buried	L	buried
442	F	exposed	F	exposed	F	exposed	I	exposed
443	Y	exposed	Y	exposed	Y	exposed	Y	exposed
444	Q	exposed	V	exposed	Y	exposed	Y	exposed
445	H	exposed	R	exposed	H	buried	H	buried
446	K	exposed	N	exposed	K	exposed	K	exposed

Table 16: Key residues of E2-CD81 interaction

1a - CD 81 Complex		1b-CD81 Complex		3a-CD81 Complex		3b-CD81 Complex	
CD81	**1a**	**CD81**	**1b**	**CD81**	**3a**	**CD81**	**3b**
ASP117	ARG455	ASP128	**GLN409**	ASP117	PRO461	**ASP117**	SER419
ASP155	*ILE472*	GLN132	**LYS410**	**LYS121**	*ILE462*	CYS157	LEU441
THR163	*LEU480*	**ASP137**	*ARG483*	**LYS124**	PHE465	SER160	TYR443
ILE182	**ARG483**	SER160	PRO484	GLN132	GLN467	THR163	SER513
LEU185	TRP487	THR161	*TYR485*	SER159	GLY468	*ALA164*	*VAL516*
PHE186	HIS488	THR163	TRP487	SER160	GLY574	LEU165	VAL616
GLU188	TYR489	*ALA164*	TYR489	**THR161**	**HIS580**	**THR167**	TYR618
ASP189	**ARG587**	*LEU165*	PRO545	LEU162	SER583	SER168	*PRO619*
GLN192	**LYS588**	**THR167**		THR163	PRO588	*ILE181*	TYR620
ASP195	TYR594	ASN184		*ALA164*	ASP590	ASN184	**ARG621**
ASP196	*SER686*	*LEU185*		LEU165	PRO596	PHE186	ILE633
TYR203	GLY688	**GLU188**		THR166	GLU597	**GLU188**	**LYS635**
		ASP189		THR167	PRO682	**ASP189**	**ARG646**
		ASP195		SER168	*CYS683*	**ASP195**	
		ASP196		***LYS171***	SER684	**ASP196**	
				ASN173	PHE685	*TYR203*	
				ILE181	THR686		
				ASP196	HIS699		
					GLN700		

Note: Residues with Interaction Energy (IE) ≥ 10 kcal/mol were reported here, residues with **IE ≥ 25** kcal/mol are in **Bold**, a. a. involved in HB underlined, residues participating in *hydrophobic contacts italicized*, residues common in all four complexes are shown in red and in three in blue.

Table 17: Binding affinity of HCV protein E2-CD81 Complexes

CD81 Complex with	Non-bonded interactions				ΔSASA (Å^2)	Binding Free Energy (kcal/mol)
	E	SB	HB	H		
1a_E2	2	-	17	9	2247.0	-92.97
1b_E2	-	1	24	11	2135.2	-106.53
3a_E2	-	-	34	10	3521.1	-238.96
3b_E2	1	1	27	9	2776.3	-153.79

Note: E – Electrostatic, SB - Salt Bridge, HB - Hydrogen Bond, H – Hydrophobic

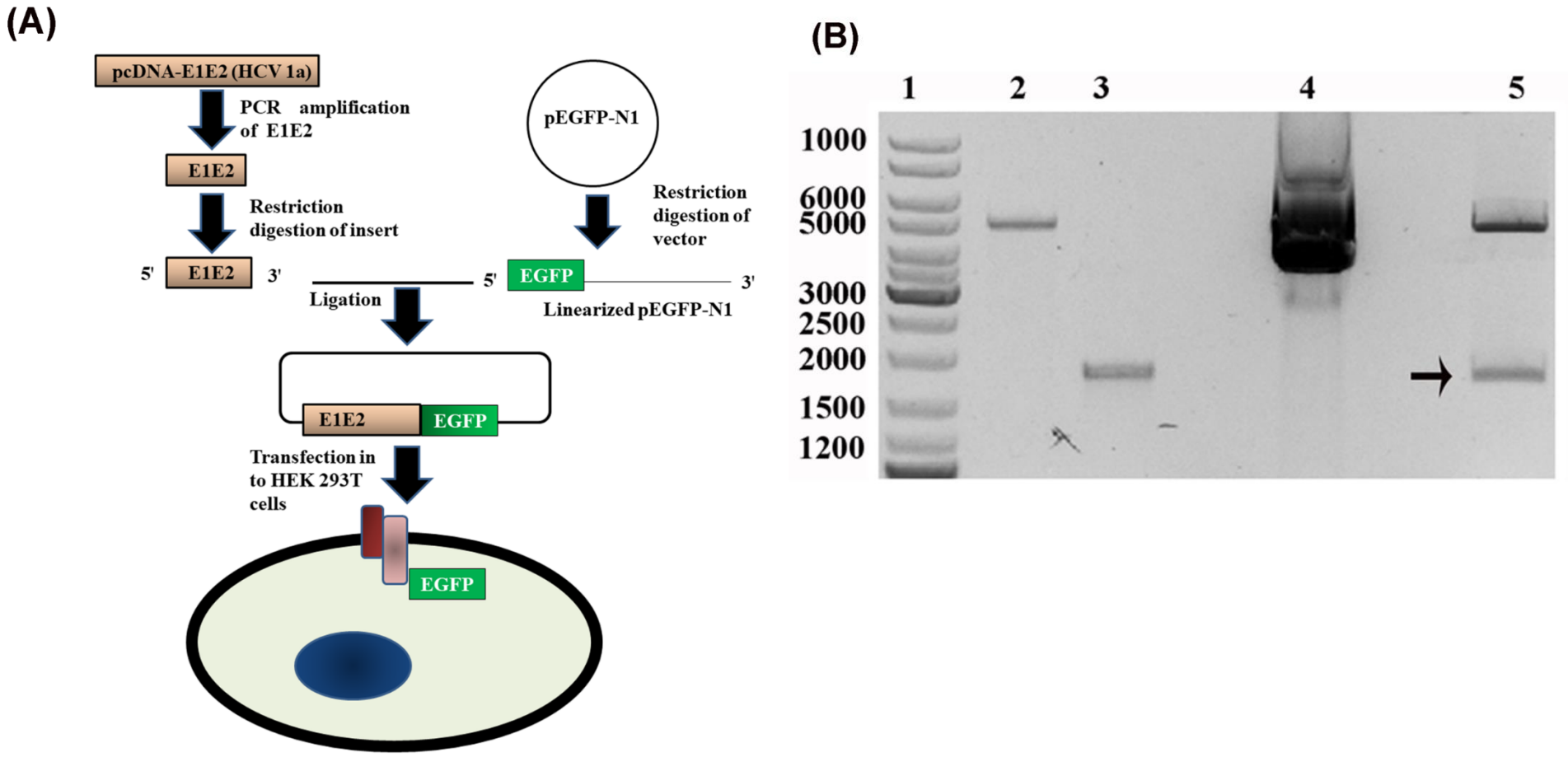


Fig. 4. Strategy of molecular subcloning of E1E2 coding region from pcDNA 3.1-E1E2 to pEGFP-N1 (A) E1E2 was PCR amplified, digested with restriction enzymes (EcoR1 and HindIII) and ligated into pEGFP-N1. Transfection of pE1E1-EGFP in HEK 293T cells was expected to yield a fluorescent protein localized on membranes. (B) Confirmation of a positive clone of pE1E2-EGFP by restriction digestion with EcoRI and HindIII. Lane1: marker, Lane2: digested empty vector (pEGFPN1), Lane3: digested PCR product (E1E2), lane4: uncut vector (pEGFP-N1) and lane5: digested clone of EGFPN1-E1E2 indicating insert drop out (arrow).

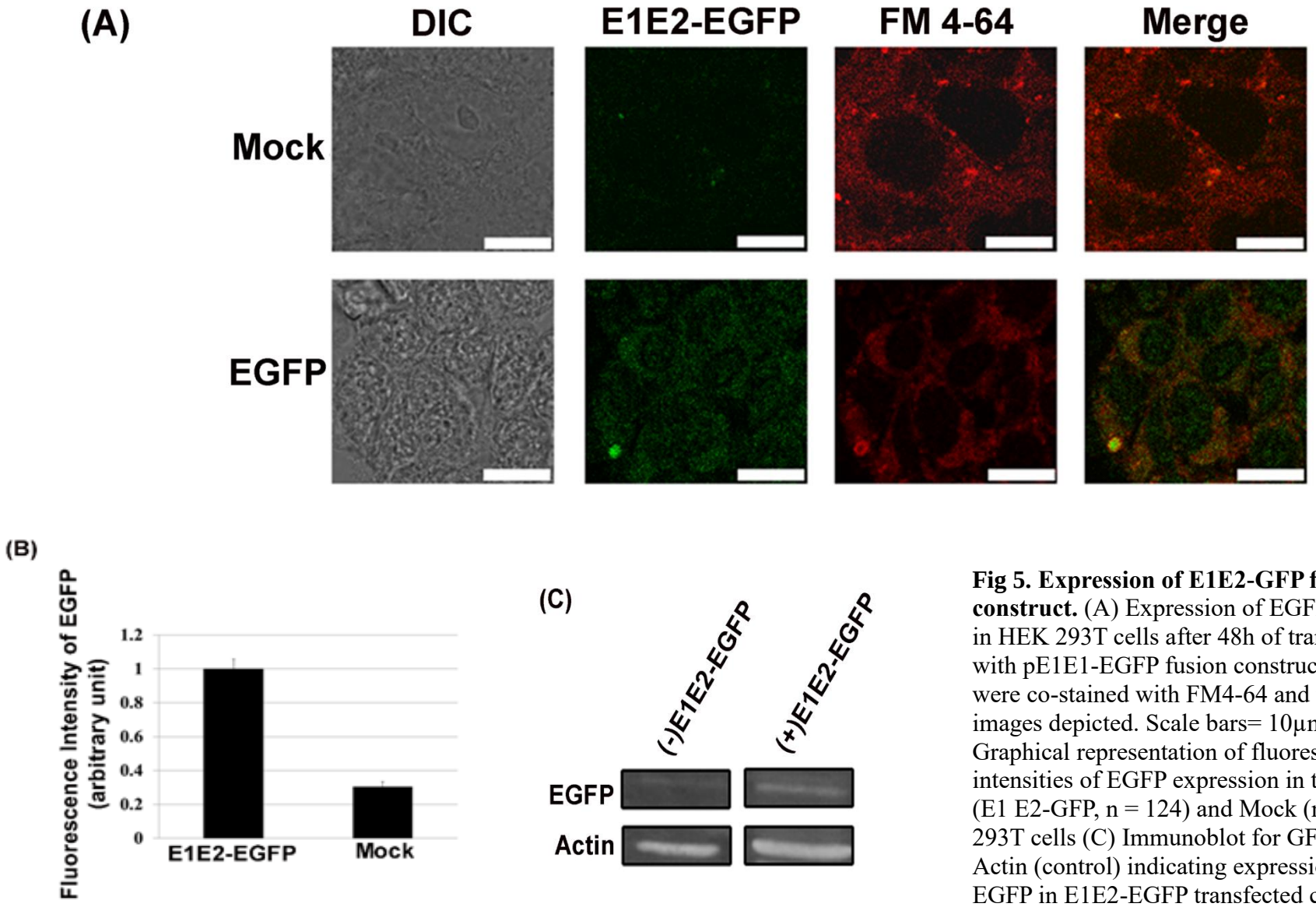


Fig 5. Expression of E1E2-GFP fusion construct. (A) Expression of EGFP (green) in HEK 293T cells after 48h of transfection with pE1E1-EGFP fusion construct. Cells were co-stained with FM4-64 and the merged images depicted. Scale bars= 10µm. (B) Graphical representation of fluorescence intensities of EGFP expression in transfected (E1 E2-GFP, n = 124) and Mock (n=68) in 293T cells (C) Immunoblot for GFP and Actin (control) indicating expression of EGFP in E1E2-EGFP transfected cell lysate.

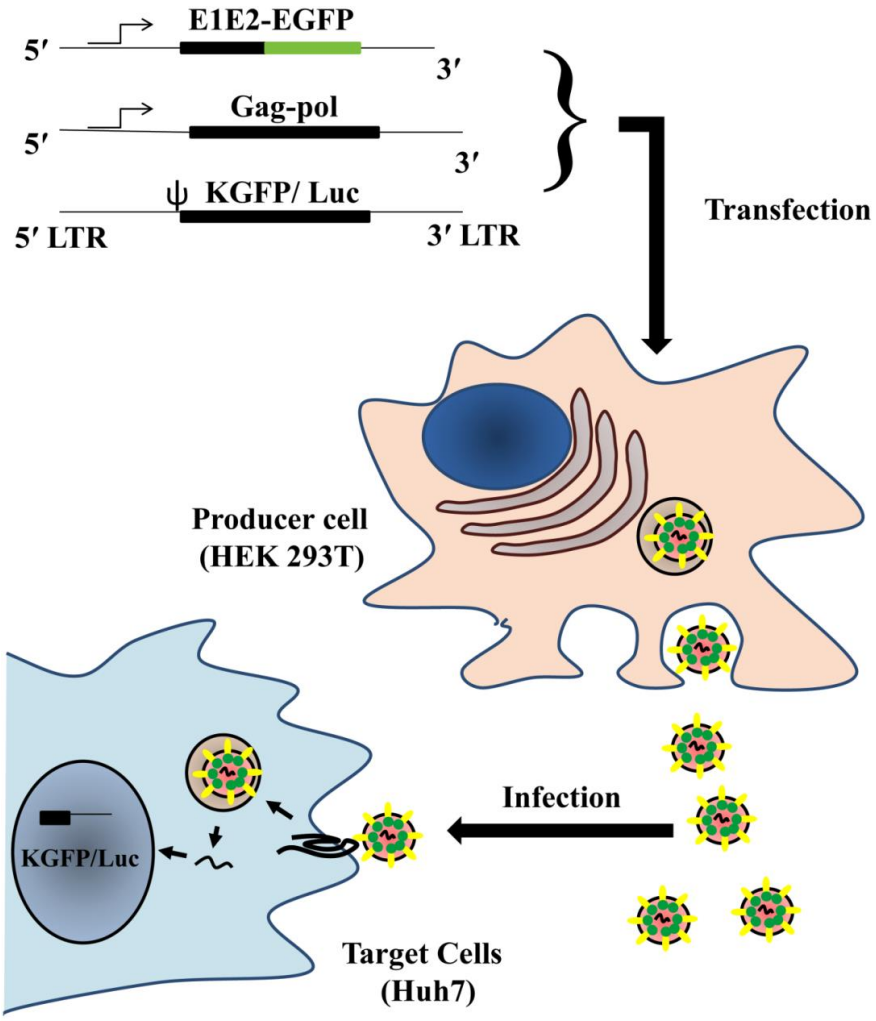


Fig 6. Production of GFP-expressing HCV pseudoparticles (HCVpp). Human embryonic kidney cells (HEK 293T) were transfected with three expression vectors; E1E2-GFP, Gag/Pol packaging construct and KGFP or Luciferase reporter construct. Fluorescent pseudoparticles extracted from the media of 293T cells were used for infection / entry studies in Huh7 cells.

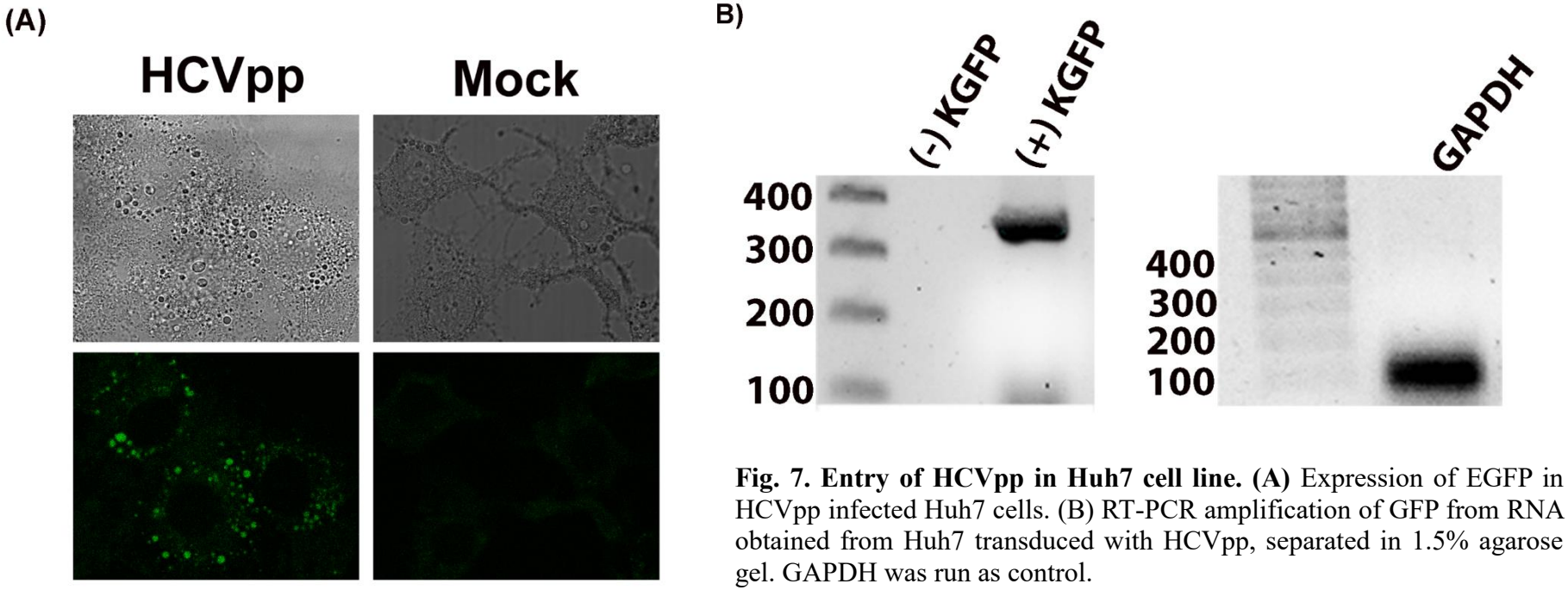


Fig. 7. Entry of HCVpp in Huh7 cell line. (A) Expression of EGFP in HCVpp infected Huh7 cells. (B) RT-PCR amplification of GFP from RNA obtained from Huh7 transduced with HCVpp, separated in 1.5% agarose gel. GAPDH was run as control.

Table 18. Transduction efficiency of HCVpp in Huh7 cells

Total number of cells (N)	**Number of Transduced cells (N)**	**Transduction efficiency (%)**
1121	729	71

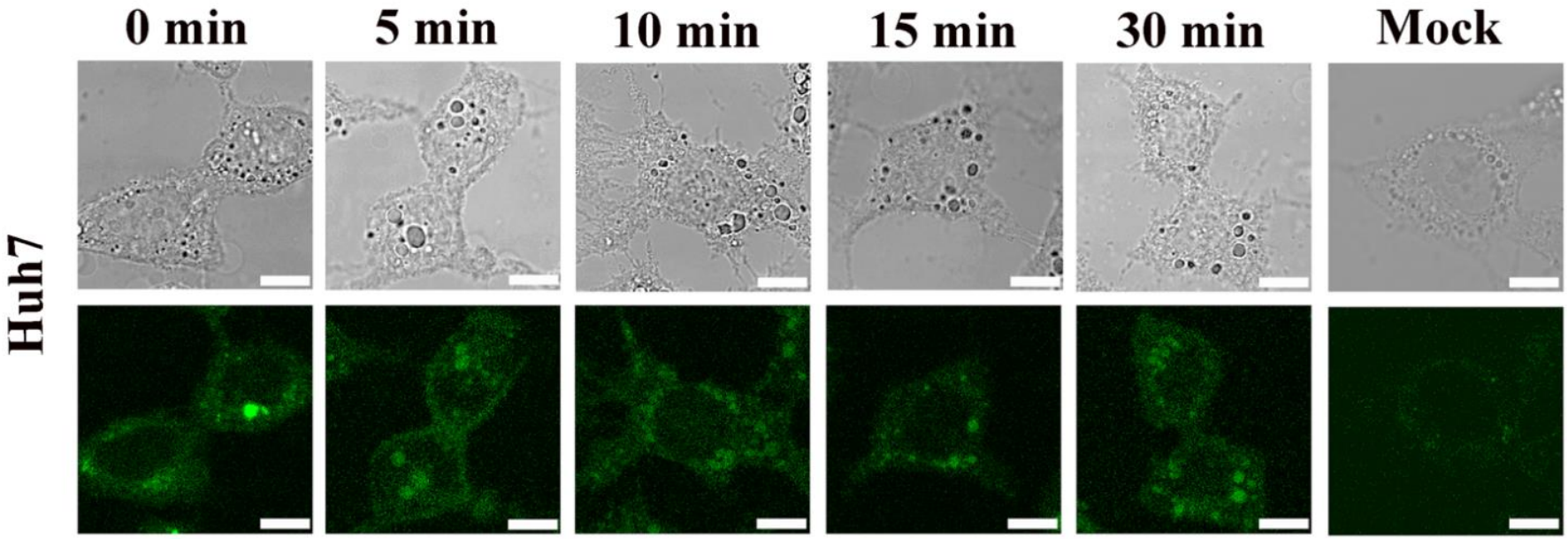
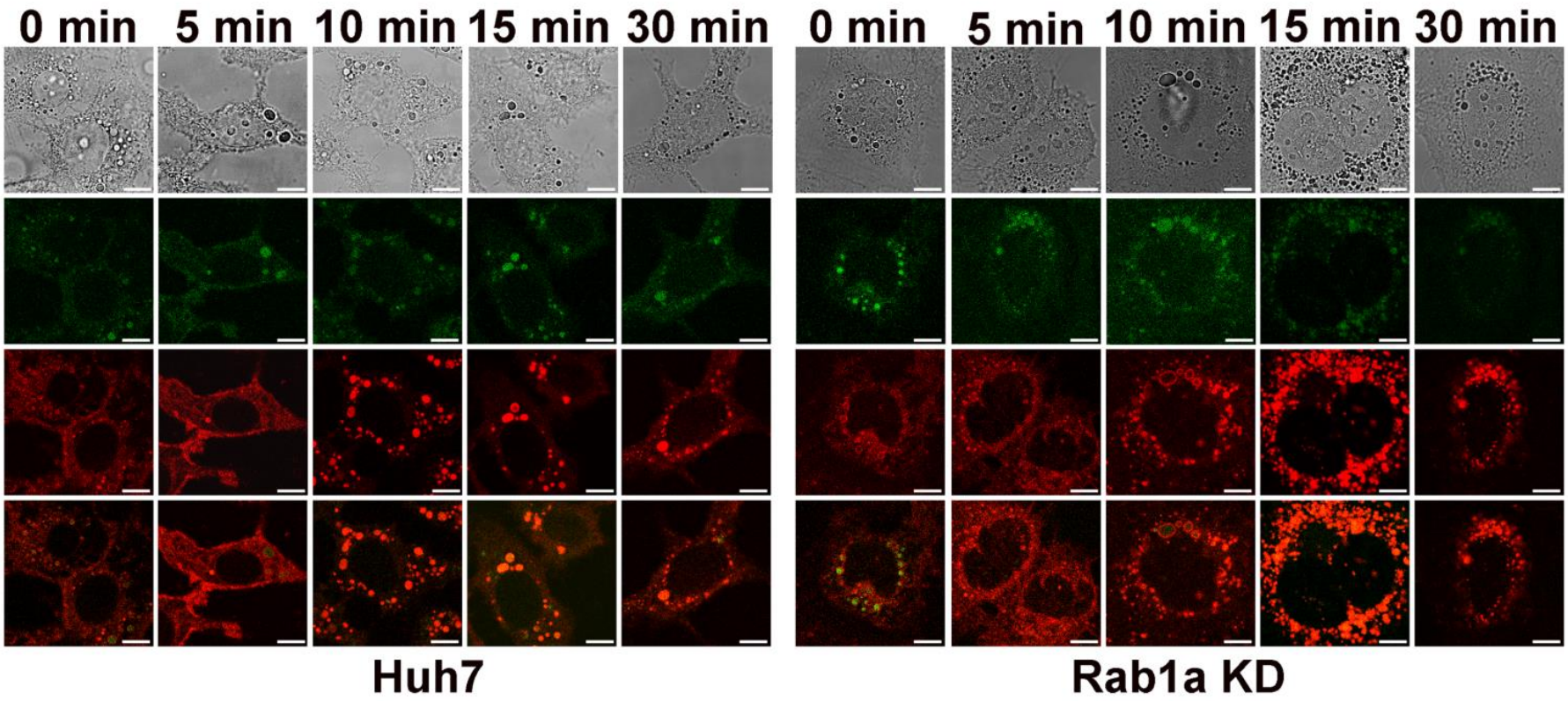


Fig. 8. HCV pseudoparticles infection and trafficking in Huh7 cells. Images of HCVpp transduced in Huh7 cells, fixed at indicated times of post internalisation and analysed via confocal microscope. Scale bars= 10 μm.

(A)


0 min
5 min
10 min
15 min
30 min
0 min
5 min
10 min
15 min
30 min
Huh7
Rab1a KD

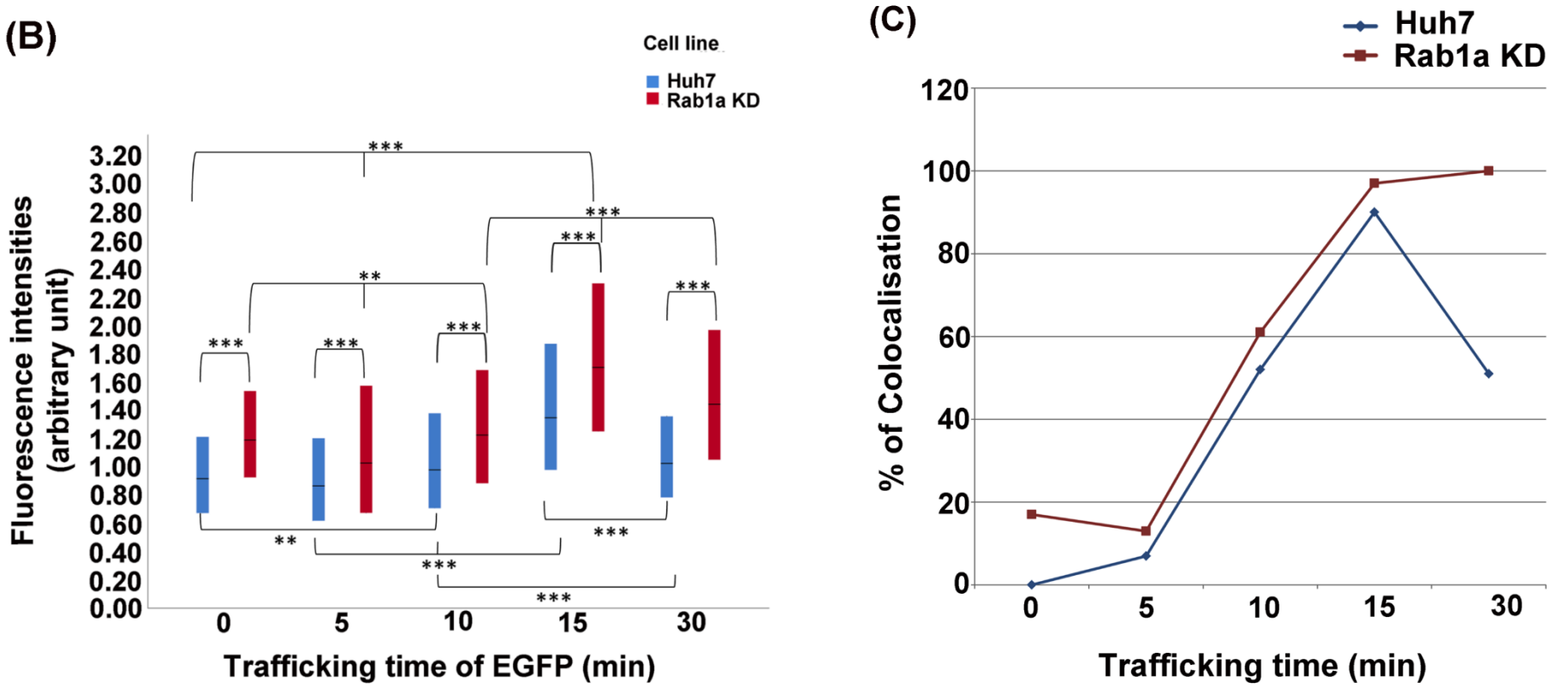


Fig. 9. Time lapse trafficking of HCVpp in Rab1a KD cell line. (A) Representative images of E1E1 GFP (green) trafficking in Huh7 (left) and Rab1a KD (right) cells for indicated time points. Cells were co-stained with FM 4-64 indicated in red. Top panel DIC images, second panel HCVpp trafficking depicted in green. Third panel: FM4-64 staining and bottom panel depicting merged images of HCVpp and FM4-64. Scale Bars= 10µm. (B) Integrated density (IntDen) of EGFP from each of the individual infected cells (Huh7 and Rab1aKD) are represented in a box plot profile. Each box represents that median, first and third quartile. ** represents significant difference ($p < 0.05$) and *** represents significant difference ($p < 0.001$) as tested by ANOVA. IntDen of images for each of the time point were counted as N. For Huh7 (0 min), N= 1748; for 5min, N= 1018; for 10min, N= 768; for 15min, N= 556; for 30min, N= 906 and for Rab1a KD (0min), N=504; for 5min, N=286; for 10min N= 468; for 15min, N= 425; for 30min N= 555.(C) Graphical representation of positive colocalisation percentage (%) between GFP and FM 4-64 are observed at indicated time points. Total numbers of cells for each time points are counted as N. For Huh7 cell lines, N (0 min)= 67, N (5 min)= 46, N(10 min)= 33, N(15 min)= 60 and N(30min)= 76; and for Rab1a KD, N(0 min)= 30, N(5min)= 47, N(10 min)= 28, N(15 min)= 74 and N(30 min)= 30. One way ANOVA followed by Bonferonni post hoc analysis was performed for statistical analysis. t-score was determined to check the significance of Pearson's R (above threshold) values.

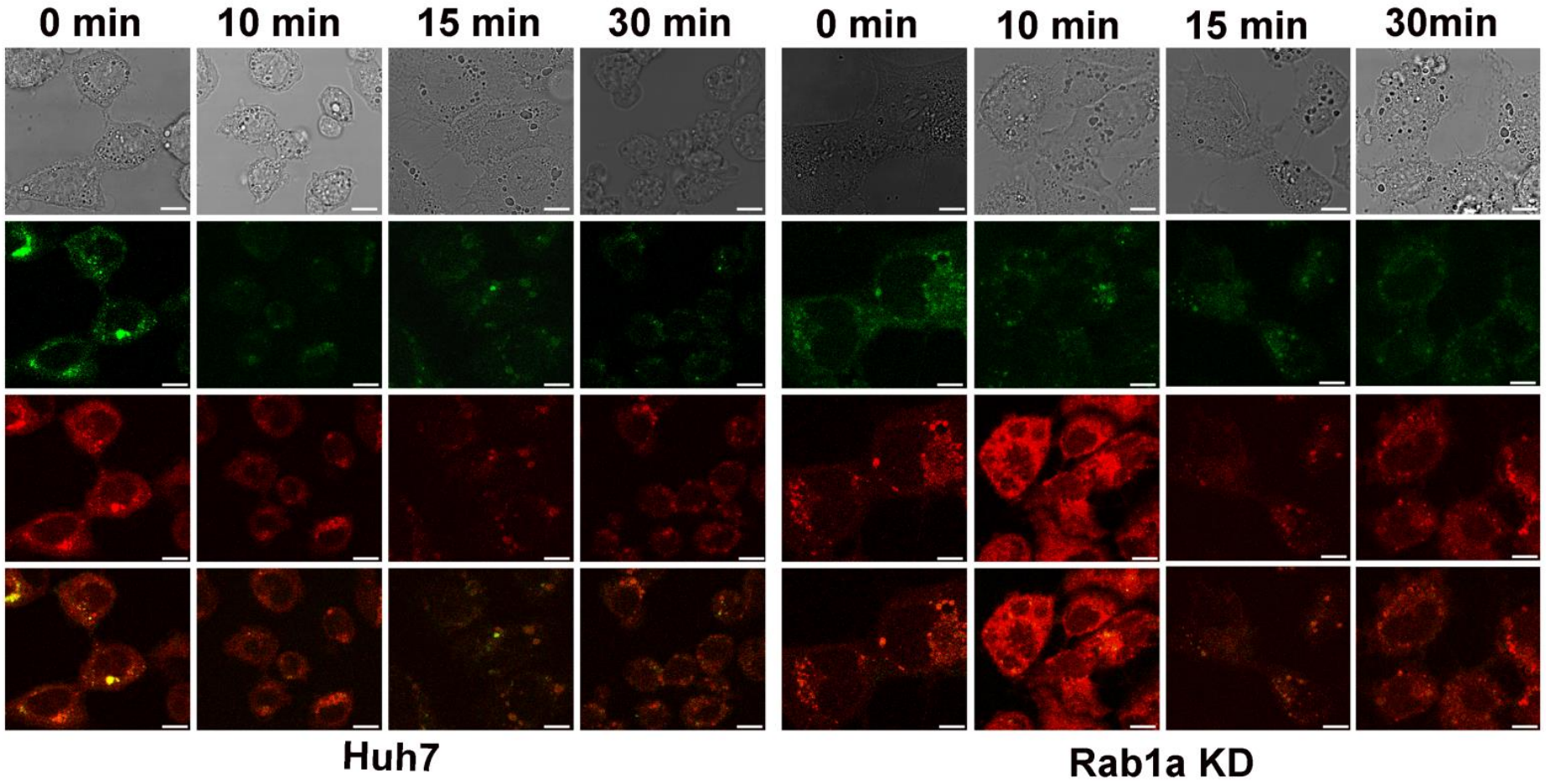


Fig. 10. HCVpp fail to reach lysosomes. HCVpp infected Huh7 and Rab1a KD cells were co-stained with lysosome marker, Lysotracker (depicted in red). Colocalisation of HCVpp (green) and Lysotracker have assayed. Total numbers of cells for each time points are counted as N. For Huh7 cell lines, N (0 min)= 50, N (5 min)= 45, N(10 min)= 45, N(15 min)= 50 and N(30min)= 55; and for Rab1a KD, N(0 min)= 50, N(5min)= 37, N(10 min)= 40, N(15 min)= 55 and N(30 min)= 56.

Table 19. Conventional Indian medicinal plants used in our study.

Name of plants	Common name	Plant type	Family	Portion Used	Reference
Psidium guajava L.	Guava	Shrub/ Small tree	Myrtaceae	leaf	[1]
Magnolia champaca L.	Champa	Large evergreen tree	Magnoliaceae	leaf	[2]
Syzygium cumini L.	Black Plum	Evergreen tropical tree	Myrtaceae	leaf	[3]
Tamarindus indica L.	Tamarind	Leguminous tree	Fabaceae	leaf	[4]

1. Dakappa SS, Adhikari R, Timilsina SS, Sajjekhan S. A REVIEW ON THE MEDICINAL PLANT PSIDIUM GUAJAVA LINN. (MYRTACEAE). *J Drug Deliv Ther*. 2013;3(2). doi:10.22270/jddt.v3i2.404
2. Lipp FJ, Perry LM, Metzger J. Medicinal Plants of East and Southeast Asia: Attributed Properties and Uses. *Brittonia*. 1981;33(2). doi:10.2307/2806338
3. Chaudhari B. and MK. Syzygium cumini (L.) SKEELS: A POTENTIAL SOURCE OF NUTRACEUTICALS. *Int J Pharm Biol Sci*. 2012;2(1).
4. Martinello F, Soares SM, Franco JJ, et al. Hypolipemic and antioxidant activities from Tamarindus indica L. pulp fruit extract in hypercholesterolemic hamsters. *Food Chem Toxicol*. 2006;44(6). doi:10.1016/j.fct.2005.10.011

Table 20. Compounds found in *Guava, Champa, Black Plum and tamarind* leaves and their pharmacological effect.

Plants name	**Active compound**	**ActivityKnown**	Referenc es
Guava	Kaempferol, flavonoids, gallic acids, cetechin, rutin, naringenin.	Polyphenol antioxidant, antiviral agent, antioxidant, anti-inflammatory, antibacterial, antifungal, anti-allergic, hepatoprotctive, anti-cholesterol activity.	1, 2, 3, 4, 5, 6, 7
Champa	Tromethamine, benzoic acid, amphetamines, andrographolide.	Analgesic and anti-inflammatory activity; antibacterial, antifungal and antiviral activity; CNS stimulants; antiallergic, hepatoprotective activity.	8, 9
Black Plum	Quercetin, myricetin, mycaminose, β-sitosterol, tannic acid, ellagic acid, caffeic acid, ferulic acid, epicatechin.	Anti-inflammatory, antioxidant, antiviral activity, antimicrobial activity, reduce cholesterol levels, anti-inflammatory agent, anti-invasive, anti-metastatic.	10, 11, 12
Tamarind	Limonene, pentadecanol, linalool, α-Pinene	Antiviral agent, antibacterial activity, anti-inflammatory, anti-oxidant, antifungal agents.	13, 14

Footnote:

1. Chen KC, Peng CC, Chiu WT, et al. Action mechanism and signal pathways of psidium guajava L. aqueous extract in killing prostate cancer LNCaP cells. *Nutr Cancer*. 2010;62(2). doi:10.1080/01635580903407130
2. Sriwilaijaroen N, Fukumoto S, Kumagai K, et al. Antiviral effects of Psidium guajava Linn. (guava) tea on the growth of clinical isolated H1N1 viruses: Its role in viral hemagglutination and neuraminidase inhibition. *Antiviral Res*. 2012;94(2). doi:10.1016/j.antiviral.2012.02.013
3. Metwally AM, Omar AA, Harraz FM, El Sohafy SM. Phytochemical investigation and antimicrobial activity of Psidium guajava L. leaves. *Pharmacogn Mag*. 2010;6(23). doi:10.4103/0973-1296.66939
4. Ryu NH, Park KR, Kim SM, et al. A hexane fraction of guava leaves (Psidium guajava L.) induces anticancer activity by suppressing AKT/mammalian target of rapamycin/ribosomal p70 S6 kinase in human prostate cancer cells. *J Med Food*. 2012. doi:10.1089/jmf.2011.1701
5. Roy CK, Kamath J V., Asad M. Hepatoprotective activity of Psidium guajava Linn. leaf extract. *Indian J Exp Biol*. 2006;44(4).
6. Nair R, Chanda S. In-vitro antimicrobial activity of Psidium guajava L. leaf extracts against clinically important pathogenic microbial strains. *Brazilian J Microbiol*. 2007;38(3). doi:10.1590/S1517-83822007000300013

7. Chen KC, Peng CC, Peng CH, Hsieh CL, Peng RY. The aqueous soluble polyphenolic fraction of Psidium guajava leaves exhibits potent anti-angiogenesis and anti-migration actions on DU145 cells. *Evidence-based Complement Altern Med*. 2011;2011. doi:10.1093/ecam/neq005
8. Khan MR, Kihara M, Omoloso AD. Antimicrobial activity of Michelia champaca. *Fitoterapia*. 2002. doi:10.1016/S0367-326X(02)00248-4
9. RUWALI P, ADHIKARI M, SHARMA S. PHYTOCHEMICAL AND ANTIOXIDANT PROPERTIES OF VARIOUS EXTRACTS OF MICHELIA CHAMPACA LEAVES. *Int J Pharm Pharm Sci*. 2019. doi:10.22159/ijpps.2019v11i5.31745
10. Arya SS, Pegu K, Sadawarte PD. Bioactive Compounds and Health Benefits of Jamun (Syzygium cumini). In: ; 2018. doi:10.1007/978-3-319-54528-8_56-1
11. Balyan U, Verma SP, Sarkar B. Phenolic compounds from Syzygium cumini (L.) Skeels leaves: Extraction and membrane purification. *J Appl Res Med Aromat Plants*. 2019;12. doi:10.1016/j.jarmap.2018.12.002
12. Kumar A, Anu, Mohan A, Sharma NR, Rehman H. Antibacterial, antioxidant analysis of phytochemical extracts derived from seeds of Syzygium cumini L. Against pathogenic bacteria. *Res J Pharm Technol*. 2017;10(8). doi:10.5958/0974-360X.2017.00481.4
13. Escalona-Arranz J, Péres-Roses R, Urdaneta-Laffita I, Camacho-Pozo M, Rodríguez-Amado J, Licea-Jiménez I. Antimicrobial activity of extracts from Tamarindus indica L. leaves. *Pharmacogn Mag*. 2010;6(23). doi:10.4103/0973-1296.66944
14. Antimicrobial Activity of Leaves Extracts of Tamarindus indica L. (Magyee). *Myanmar Heal Sci Res J*. 2019;31(2). doi:10.34299/mhsrj.00935

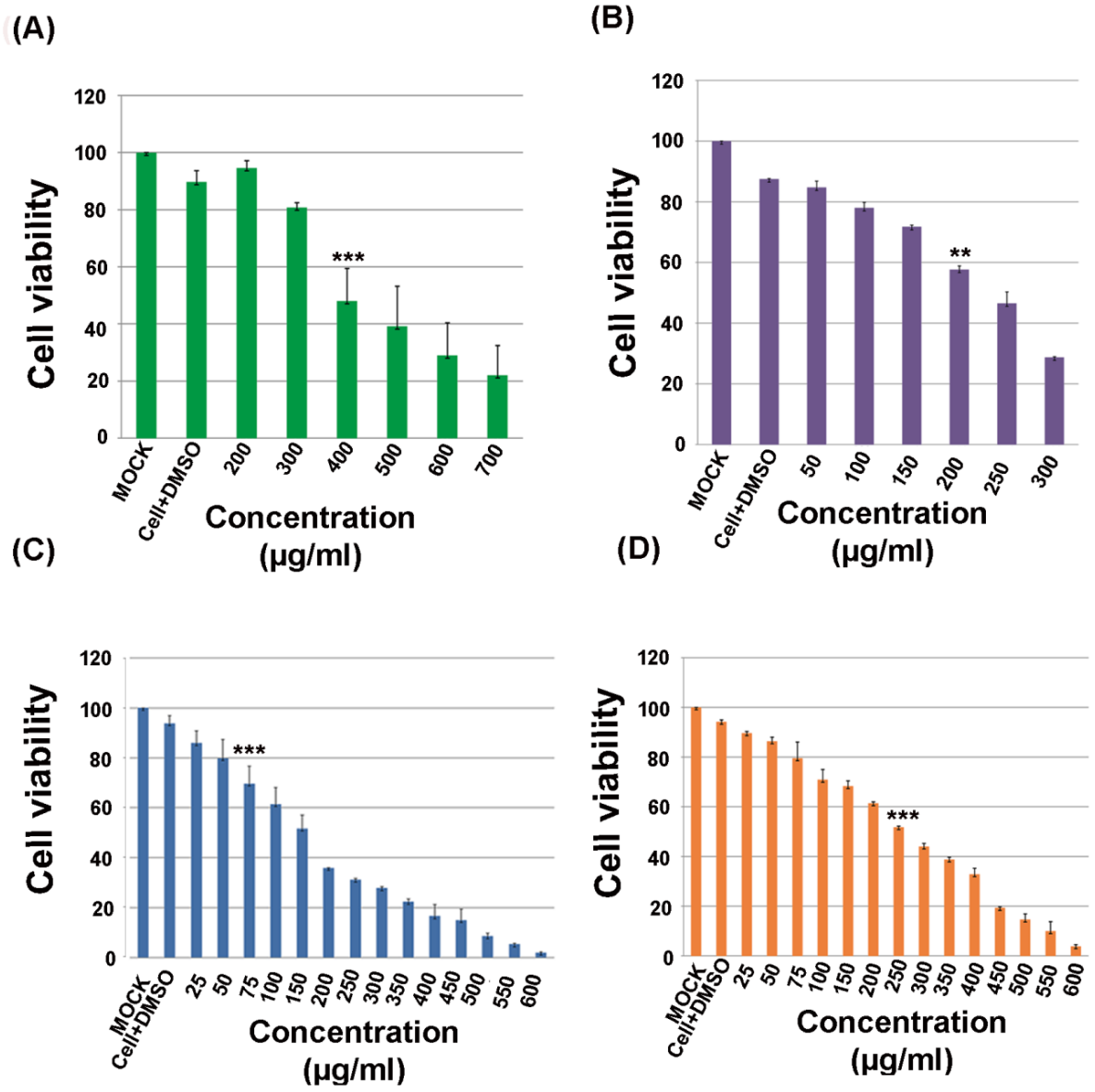


Fig. 11. MTT analysis for of the toxicity analysis of the methanolic leaf extracts. Cell viability in percentage depicted for (A) Guava, (B) Champa, (C) Black plum and (D) Tamarind leaf extracts in Huh7 cells. Errors bars indicate standard deviation of three independent experiments. *** represents significant difference of $p < 0.001$ and ** represents significant difference of $p < 0.05$. 1-tailed pared t-test was performed to determine statistical significance.

Table 21: Optimised concentration of leaf extract by MTT assay.

Methanolic Leaf Extract	Selected concentration (µg/ml)
Guava	**300**
Champa	**150**
Black plum	**50**
Tamarind	**100**

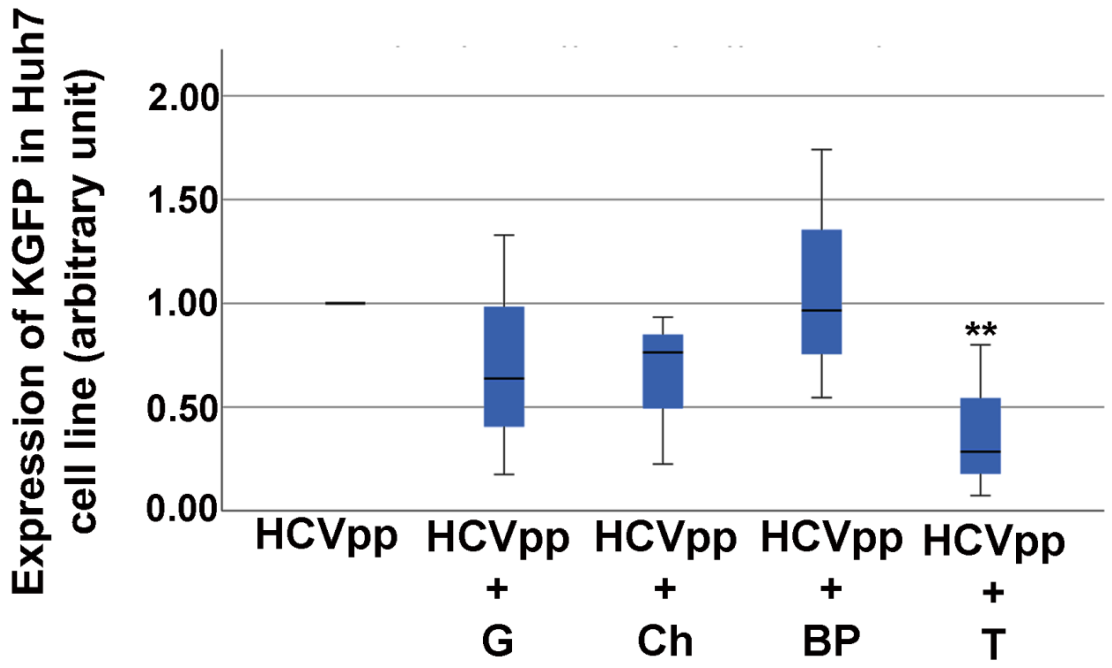


Fig. 12. Presence of KGFP in HCVpp infected Huh7 cells indicate HCV entry inhibition by plant extracts. G indicate Guava leaf extract, Ch: Champa leaf extract, BP: Black plum leaf extract and T: Tamarind leaf extracts. KGFP expressions are represented in a box plot profile. Each box represents the median, first and third quartile. Error bars indicate standard deviation, ** represents significant difference ($p < 0.05$) as tested by one way ANOVA followed by Post hoc Tukey (Tukey's HSD) analysis.

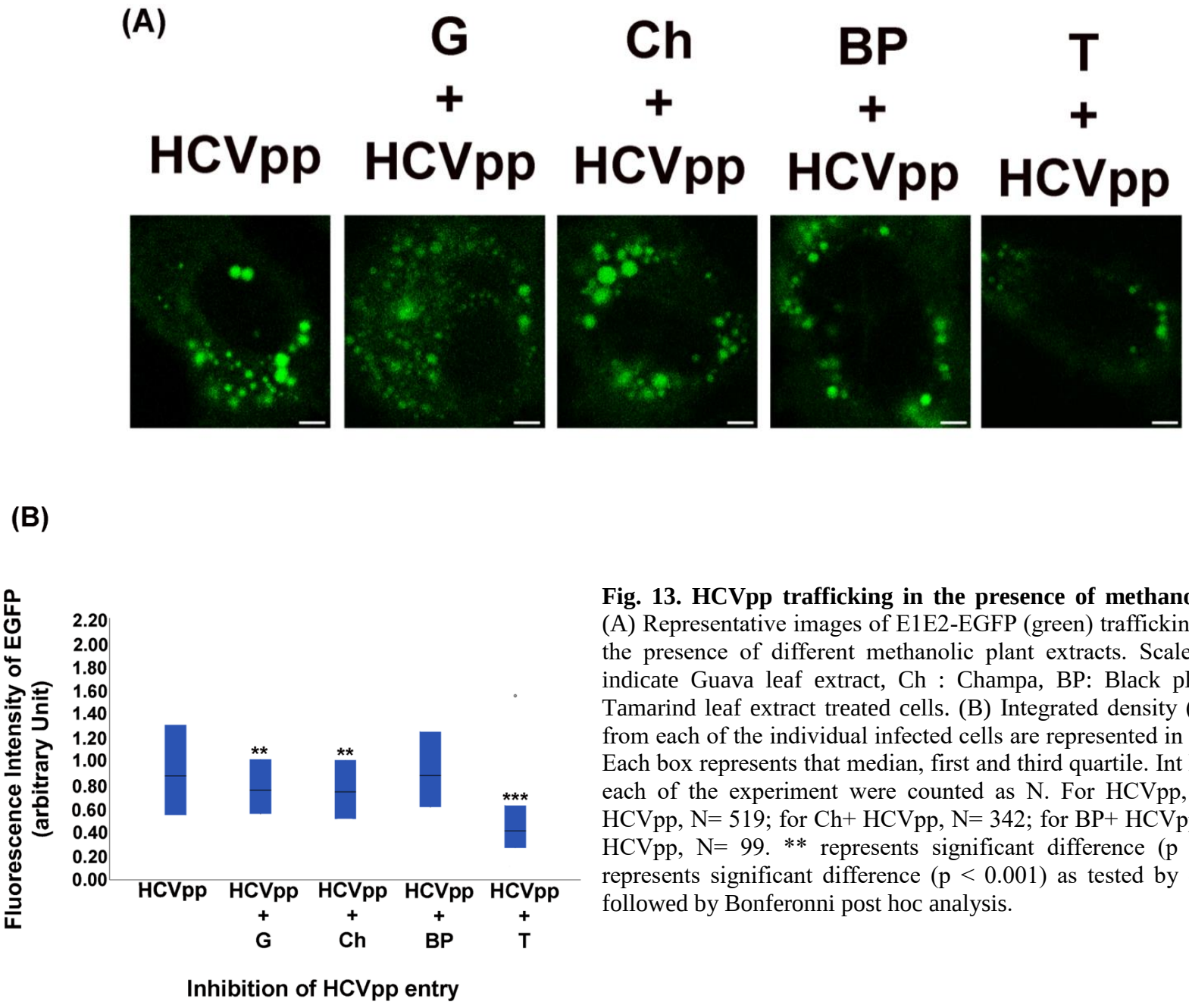
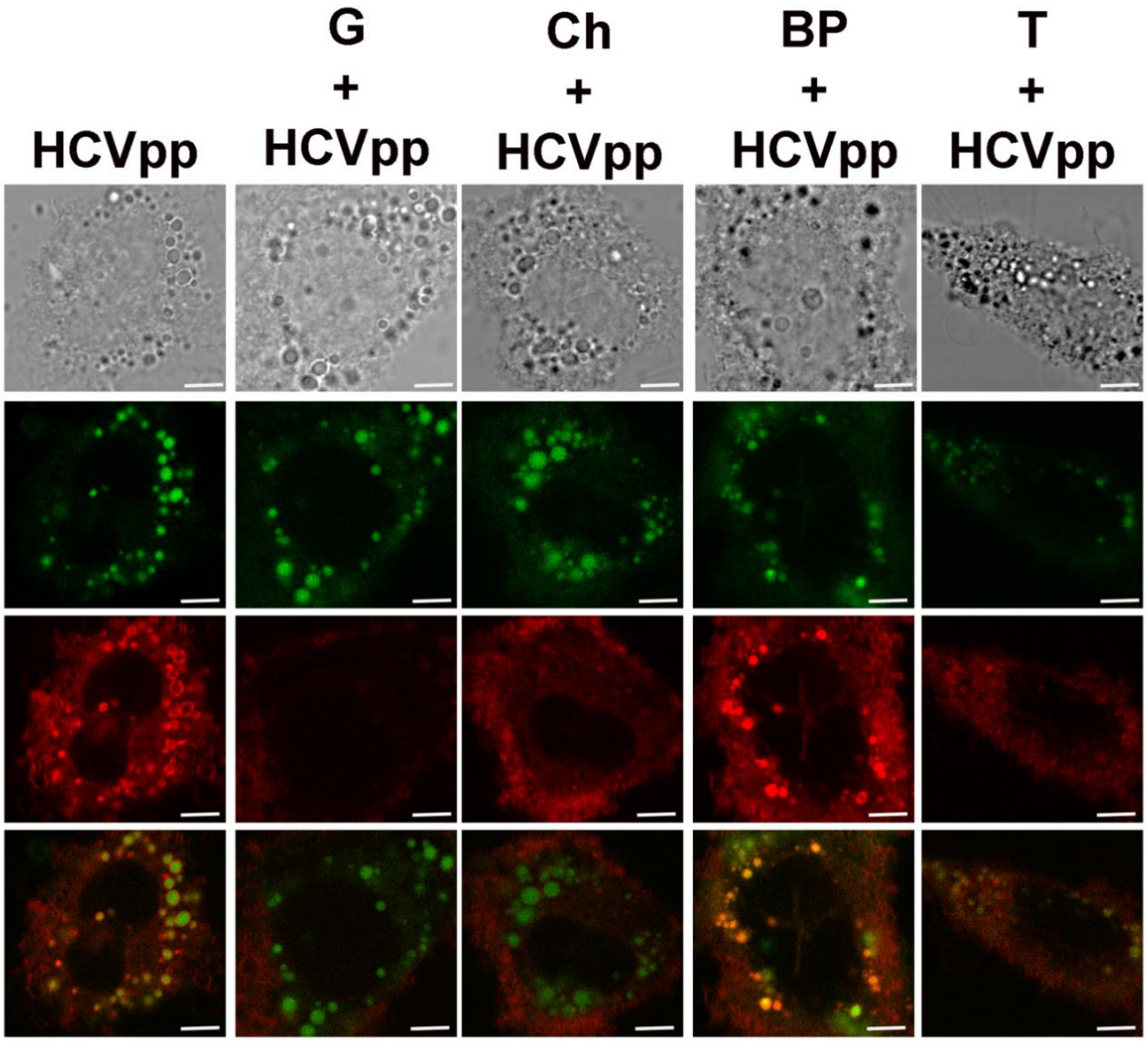


Fig. 13. HCVpp trafficking in the presence of methanolic leaf extracts. (A) Representative images of E1E2-EGFP (green) trafficking in Huh7 cells in the presence of different methanolic plant extracts. Scale Bars= 10µm. G indicate Guava leaf extract, Ch : Champa, BP: Black plum and T is for Tamarind leaf extract treated cells. (B) Integrated density (IntDen) of EGFP from each of the individual infected cells are represented in a box plot profile. Each box represents that median, first and third quartile. Int Den of images for each of the experiment were counted as N. For HCVpp, N= 303; for G+ HCVpp, N= 519; for Ch+ HCVpp, N= 342; for BP+ HCVpp, N= 121; for T+ HCVpp, N= 99. ** represents significant difference ($p < 0.05$) and *** represents significant difference ($p < 0.001$) as tested by one-way ANOVA followed by Bonferonni post hoc analysis.


HCVpp
G + HCVpp
Ch + HCVpp
BP + HCVpp
T + HCVpp

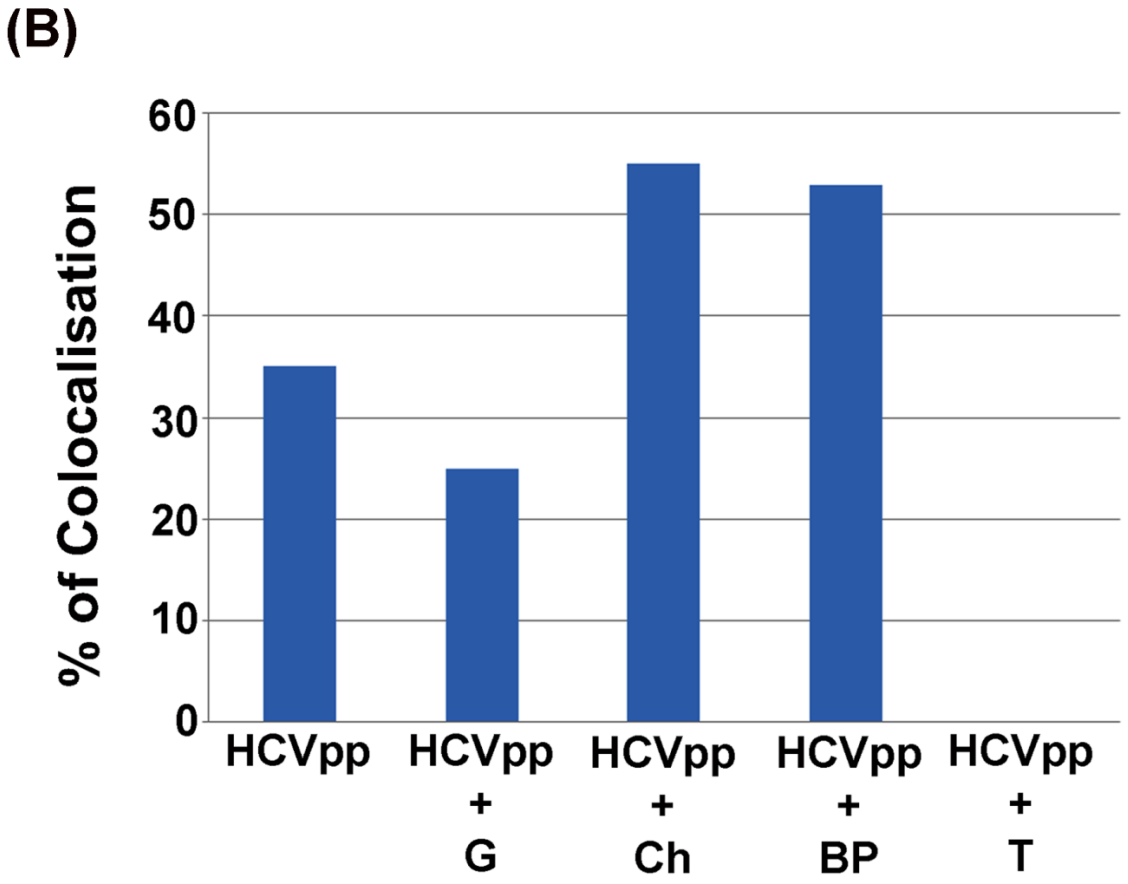
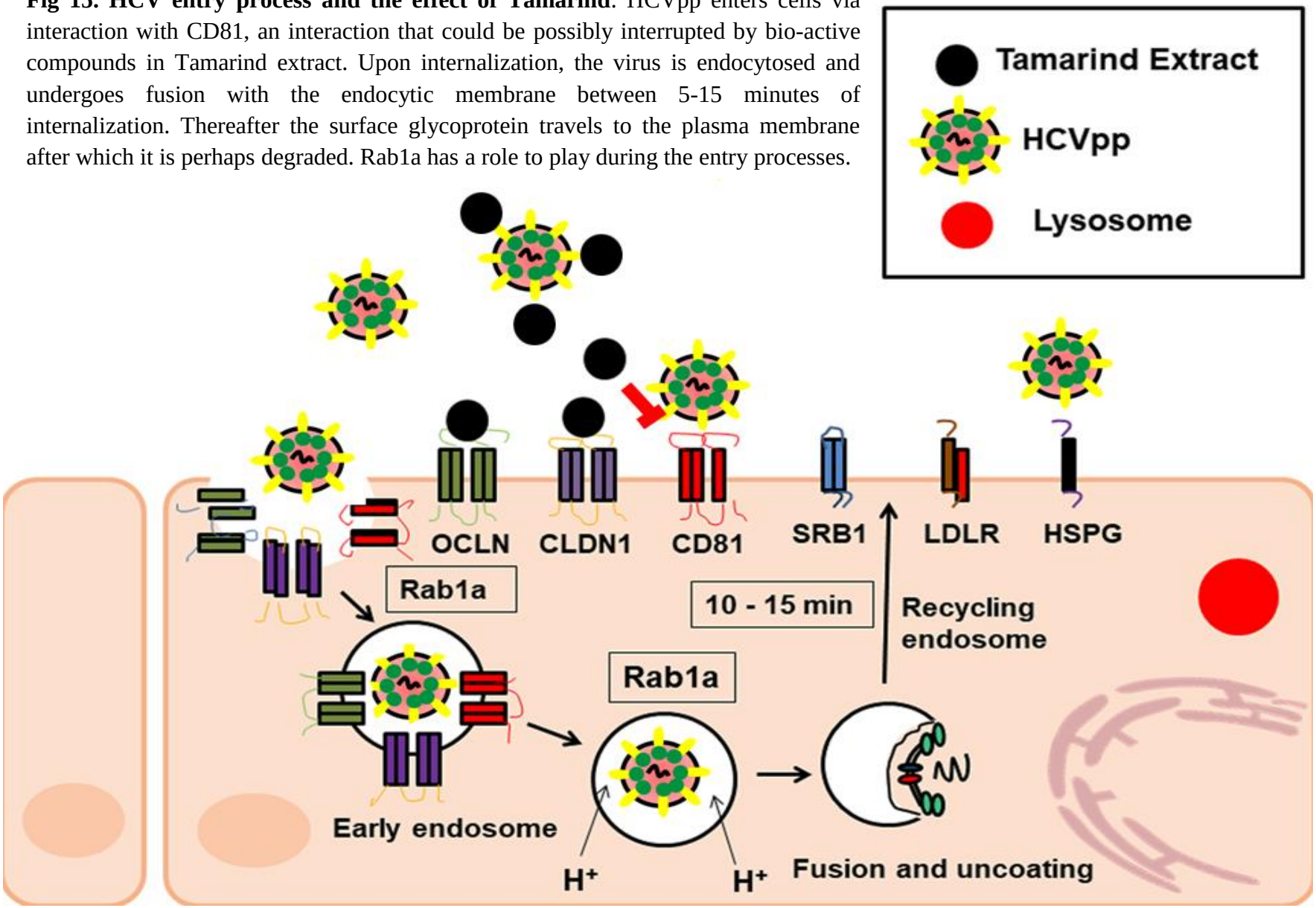


Fig 14. Colocalisation of EGFP (green) and plasma membrane marker FM 4-64 (Red). Representative images of E1E2-EGFP (green) trafficking in Huh7 cells in the presence of different methanolic leaf extracts colocalised with FM 4-64. Scale Bars= 10µm. G = Guava leaf extract, Ch = Champa, BP = Black plum and T = Tamarind leaf extract treated cells. (A) Graphical representation of positive colocalisation percentage (%) between GFP and FM 4-64 as observed in replicate experiments. Total numbers of cells for each experiment are counted as N. For HCVPP cell lines, N= 20, N (G+ HCVPP)= 20, N(Ch+ HCVPP)= 20, N(BP+ HCVPP)= 17 and N(T+ HCVPP)= 17. t-score was determined to check the significance of Pearson's R (above threshold) values.

Fig 15. HCV entry process and the effect of Tamarind. HCVpp enters cells via interaction with CD81, an interaction that could be possibly interrupted by bio-active compounds in Tamarind extract. Upon internalization, the virus is endocytosed and undergoes fusion with the endocytic membrane between 5-15 minutes of internalization. Thereafter the surface glycoprotein travels to the plasma membrane after which it is perhaps degraded. Rab1a has a role to play during the entry processes.



9. Appendix 1

9.1 Chemicals and reagents used in the study

Serial Number	Chemicals / Reagents	purchased from	Used for
1	Dulbecco's modified Eagle's medium (DMEM) media	Thermo Fisher Scientific, USA	Culture medium for HEK 293T cell line.
2	Roswell Park Memorial Institute (RPMI) media	Thermo Fisher Scientific, USA	Culture medium for Huh7 cell line.
3	Fetal bovine Serum	Thermo Fisher Scientific, USA	Provide growth factors for cell survival and proliferation of cultured cells.
4	penicillin-streptomycin	Thermo Fisher Scientific, USA	Antibiotic souse of cell culture medium.
5	puromycin	Thermo Fisher Scientific, USA	Used as a antibiotic souse for Rab1a KD stable cell line.
6	Phosphate-buffered saline (PBS)	Sigma Aldrich, Germany	Buffer solution
7	Hank's Balanced Salt Solution (HBSS)	Sigma Aldrich, Germany	Buffer solution
8	Rabbit anti-GFP antibody (G1544)	Sigma Aldrich, Germany	Detect EGFP protein by western blot.
9	FM 4-64 lipophilic	Thermo Fisher Scientific, USA (Molecular probes, Invitrogen detection technologies).	Lipophilic dye, act as a plasma membrane marker.
10	Lyso Tracker	Thermo Fisher Scientific, USA (Molecular Probes, Life Technologies).	Act as a lyso-sensor molecular probe.

9.2 Different cell lines used in our study.

Serial Number	Cell Line	Source	Culture Conditions	Used for
1	HEK 293T Cells	American Type Culture Collection (ATCC).	Grown in DMEM, containing L-glutamine, supplemented with 10% FBS and 1% penicillin streptomycin.	Used for transfection of foreign genes for overexpression.
2	Huh7 Cells	American Type Culture Collection (ATCC).	Grown in RPMI, containing L-glutamine, supplemented with 10% FBS and 1% penicillin streptomycin.	This hepatocarcinoma cell line is used for Hepatitis C virus transduction.
3	Rab1a Knock Down (Rab1a KD) Huh7 Cells.	As a kind gift from Dr. Allan Wolkoff, Albert Einstein College of Medicine, Bronx, NY.	Grown in RPMI, containing L-glutamine, supplemented with 10% FBS and 1% penicillin streptomycin. Additionally supplement of 2.0 µg/ml Puromycin.	This Rab1a KD hepatocarcinoma Cell line used to analyse Rab1a trafficking.

9.3 Vectors used in our study.

Plasmid	Source	Used for
pcDNA-E1E2	Jean Dubuisson, Université Ede Lille, France.	Containing the full length cDNA of HCV genotype 1a (H77) strain.
pTG5349	Jean Dubuisson, Université Ede Lille, France.	Containing retroviral structural genes Gag-Pol.
pLenti-PGKGFP	Dr. Jayanta Roychoudhury , Albert Einstein College of Medicine, New York.	Containing KGFP reporter gene.

pEGFP-N1	Clontech laboratories, USA.	Containing EGFP reporter gene.

9.4 Primers used in our study

Primers	**Sequence**
E1E2 sequence primer (forward)	5′ CGAAGCTTGCATGGGTTGCTCTTTC 3′.
E1E2 sequence primer (reverse)	5′ CAGAATTCCCGCCTCCGC 3′
KGFP sequence primer (forward)	5´TCGACAGTCAGCCGCATCT3´
KGFP sequence primer (reverse)	5´CCGTTGACTCCGACCTTCA3´
GAPDH sequence primer (forward)	5´TCGACAGTCAGCCGCATCT3´
GAPDH sequence primer (reverse)	5´CCGTTGACTCCGACCTTCA3

Infection, Genetics and Evolution 69 (2019) 48–60

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Research paper

Differential interaction strategies of hepatitis c virus genotypes during entry – An in silico investigation of envelope glycoprotein E2 - CD81 interaction☆

Chayan Bhattacharjee[a,1], Suman K. Nandy[b,1,2], Pratik Das[a], Aparna Mukhopadhyay[a,*]

[a] Department of Life Sciences, Presidency University, 86/1 College Street, Kolkata, 700073, India
[b] Department of Biochemistry & Biophysics, University of Kalyani, Kalyani 741235, India

ARTICLE INFO

Keywords:
HCV
Envelope glycoprotein E2
CD81
Protein docking

ABSTRACT

Hepatitis C Virus is a blood borne pathogen responsible for chronic hepatitis in more than 71 million people. Wide variations across strains and genotypes are one of the major hurdles in therapeutic development. While genotype 1 remains the most extensively studied and abundant strain, genotype 3 is more virulent and second most prevalent. This study aimed to compare differences in the glycoprotein E2 across HCV genotypes at nucleotide, protein and structural levels. Nucleotide sequences of E2 from 29 strains across genotypes 1a, 1b, 3a and 3b revealed a stark preference for C-richness which was attributed to a distinct bias for C-rich codons in genotype 1. Genotype 3 exhibited a similar preference to a lesser extent. Amino acid level comparison revealed majority of the changes at the C-terminal half of the proteins leaving the N-terminal region conspicuously conserved apart from the two hyper variable regions. Amino acid changes across genotypes were mostly polar-nonpolar alterations. In silico models of E2 glycoproteins and docking analysis with the energy minimized PDB-CD81 model revealed unique interacting residues in both E2 and CD81. While several CD81 binding residues were common for all four genotypes, number and composition of interacting residues varied. The interacting residues of E2 were however unique for each genotype. E2 of genotype 3a and CD81 had the strongest interaction. In conclusion this is the first comprehensive study comparing E2 sequences across genotypes 1a, 1b, 3a and 3b revealing stark genotype-specific differences which requires more extensive investigation.

1. Introduction

Hepatitis C virus (HCV), a member of the Flaviviridae family, is one of the leading causes of chronic hepatitis, liver cirrhosis, and hepatocellular carcinoma worldwide. HCV, an enveloped virus with an approximately 9.6 kb long, positive stranded RNA genome, encodes a single polyprotein of about 3010 amino acids. Post-translationally, this polyprotein is cleaved (Fields et al., 2013) to generate at least 11 structural and non-structural proteins. Structural proteins include two viral glycoproteins, E1 and E2 (Yagnik et al., 2000) located at the N terminus of the polyprotein (Moradpour and Penin, 2013). HCV heterodimer E1/E2, the complex responsible for viral entry and tropism, are type I transmembrane proteins with an N terminal ectodomain and C- terminal transmembrane domain (TMD). The exact role of E1 in viral entry is poorly understood, but is believed to be related to the fusion process (Boo et al., 2012; Callens et al., 2005). E2, however, is currently the better characterized subunit where it plays pivotal roles in HCV entry, i.e. interaction with entry factors like scavenger receptor BI (SR-B1) (Scarselli et al., 2002), tetraspanin CD81 (Pileri et al., 1998) along with several tight junction proteins, Claudin-1 (Evans et al., 2007) and occludin receptors (Ciesek et al., 2011). E2 is also one of the most diverse HCV proteins characterized by two hypervariable regions (HVR), HVR1 and HVR2. Previous reports have shown that both HVR1 and HVR2 are important for the interaction with the cellular receptor

Abbreviations: HCV, Hepatitis C Virus; GP, Glycoprotein; CD81, Cluster of Differentiation 81, Tetraspanin protein; E1 /E2, Envelope glycoprotein on HCV; HCVpp, HCV pseudoparticle; HCVcc, Cell culture adaptable HCV; TMD, Transmembrane domain; CUB, codon usage bias; RMS, Root Mean Square; PROVE, PROtein Volume Evaluation; SAVES, Structural Analysis and Verification Server; BLAST, Basic Local alignment Search Tool; PDB, Protein data bank; SASA, solvent accessible surface area; IE, Interaction energy; HB, hydrogen bond; VDW, Van der Waals

☆ Data in this paper is part of a doctoral thesis to be submitted in partial fulfilment of the requirements of a degree of Doctor of Philosophy from Presidency University, Kolkata.

* Corresponding author.
E-mail address: aparna.dbs@presiuniv.ac.in (A. Mukhopadhyay).

[1] Both authors contributed equally to this work.

[2] Present Address: Bioinformatics Infrastructure Facility, North-Eastern Hill University, Tura Campus, Tura, Meghalaya-794002, India.

https://doi.org/10.1016/j.meegid.2019.01.008
Received 3 August 2018; Received in revised form 12 December 2018; Accepted 8 January 2019
Available online 11 January 2019
1567-1348/ © 2019 Elsevier B.V. All rights reserved.

NITTE

Indian Virological Society & Nitte University

VIROCON 2017

26th Annual Conference of Indian Virological Society (IVS)

"Viruses to Viromes in Health and Disease"

Participation Certificate

This is to certify that ~~Prof./Dr./Mr./Ms.~~ Chayan Bhattacharjee has participated and presented ~~oral~~ / poster at the 26th Annual conference **"VIROCON - 2017"** during 7 - 9 December organised by Nitte University, Deralakatte, Mangaluru, India.

Prof. Anupam Varma
President
Indian Virological Society

Dr. G.P. Rao
General Secretary
Indian Virological Society

Prof. Iddya Karunasagar
Chairman
Organising Committee

Indian Virological Society & PGIMER, Chandigarh

INTERVIROCON 2018

27th International Conference of Virology

10th-14th November 2018

Postgraduate Institute of Medical Education and Research, Chandigarh, India

Participation Certificate

This is to certify that

Dr./Mr./Ms ..... **Chayan Bhattacharjee** .....

has participated as a Delegate / Presented Scientific Paper (~~Oral~~ / Poster) in the conference

"Global Viral Epidemics: A Challenging Threat"

Organized by the Department of Virology, Postgraduate Institute of Medical Education and Research, Chandigarh, India

from 12th - 14th November 2018.

Dr. G.P. Rao
Secretary General
Indian Virological Society

Dr. RK Ratho
Chairperson
INTERVIROCON 2018

Dr. Mini P Singh
Organizing Secretary
INTERVIROCON 2018

Dr. Kapil Goyal
Joint Organizing Secretary
INTERVIROCON 2018

HCV2019

26th International Symposium on
Hepatitis C Virus and Related Viruses
October 5-8, 2019 | COEX, Seoul, South Korea

Certificate of Attendance

Chayan Bhattacharjee

Presidency Univ., India

Attended the 26th International Symposium on
Hepatitis C Virus and Related Viruses 2019 (HCV 2019), held at
the COEX, Seoul, South Korea, from October 5-8, 2019.

October 8, 2019

Eui-Cheol Shin
KAIST, Daejeon,
South Korea

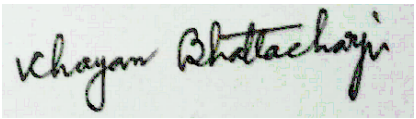
Sung Key Jang
POSTECH, Pohang,
South Korea

Kwang-Hyub Han
Yonsei Univ., Seoul,
South Korea

Takaji Wakita
National Institute of
Infectious Diseases,
Tokyo, Japan

Document Information

Analyzed document	For Plagiarism check.docx (D88186984)
Submitted	12/7/2020 12:47:00 PM
Submitted by	Aparna Mukhopadhyay
Submitter email	aparna.dbs@presiuniv.ac.in
Similarity	1%
Analysis address	aparna.dbs.presi@analysis.urkund.com

Sources included in the report

SA	**Presidency University / Irfan.docx** Document Irfan.docx (D74214607) Submitted by: aparna.dbs@presiuniv.ac.in Receiver: aparna.dbs.presi@analysis.urkund.com	1
W	URL: https://www.researchgate.net/publication/276132116_The_missing_pieces_of_the_HCV_e ... Fetched: 12/7/2020 1:19:00 PM	1
W	URL: https://worldwidescience.org/topicpages/r/rab+family+proteins.html Fetched: 5/13/2020 5:48:33 AM	1